EXTENSIBLE SPECTRALISM: REVEALING LATENT STRUCTURES IN MUSIC AUDIO FOR COMPOSITION, ANALYSIS, AND RETRIEVAL

Part I

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Musical Arts

by
Spencer Topel

May 2012

EXTENSIBLE SPECTRALISM: REVEALING LATENT STRUCTURES IN MUSIC AUDIO FOR COMPOSITION, ANALYSIS, AND RETRIEVAL

Spencer Topel, D.M.A.

Cornell University 2012

The music contained within **Part I** is an outgrowth of my residency at Cornell University, and later as a lecturer at Dartmouth College. In all but one case, the compositions predate the outcomes of my research presented in **Part II**. While some of these pieces utilize variants of spectralist techniques—such as *Incendio* (2008) and *Cyclicum* (2010)—they are not indicative of my most current creative work exploring spectral decomposition. Only *Elementary Sources* (2011) and the repertoire of compositions discussed in Chapter 5 of **Part II** point toward the extensibility of spectralism.

TABLE OF CONTENTS

SPENCER TOPEL

I N C E N D I O

for orchestra
(2008)

version 2, August 2009

Full Score

INSTRUMENTATION

duration [9' – 10']

3 Flutes
(third doubling piccolo)

2 Oboes

English Horn

2 Clarinets Bb

Bass Clarinet

2 Bassoons

Contrabassoon

4 Horns in F

2 Trumpets in C

2 Trombones

Bass Trombone

Tuba

Timpani

Percussion (3)*

Piano

Harp

Strings

*Player 1: Vibraphone

*Player 2: Glockenspiel, Sizzle Cymbal, Crash Cymbals,
Tam-Tam, Crystal Wine Glass (tuned to C#6)

*Player 3: Pipe Brake Drums 6x (tuned to octave D's, Eb's, A's)
Sizzle Cymbal, Med Susp. Cymbal, Susp. Cymbal, Bass Drum,
Crystal Wine Glass (tuned to C#6)

performance notes and glossary

Wine glasses performed mm. 176 – 183 should have a pure tone with little difference in pitch between the two glasses. A water/vinegar mixture can help produce a better tone and improve projection.

z = an uneven tremolo. This should be played with varying lengths of bow by each player individually. Violin I, Violin II, Viola, and Double Bass, mm.238 –250.

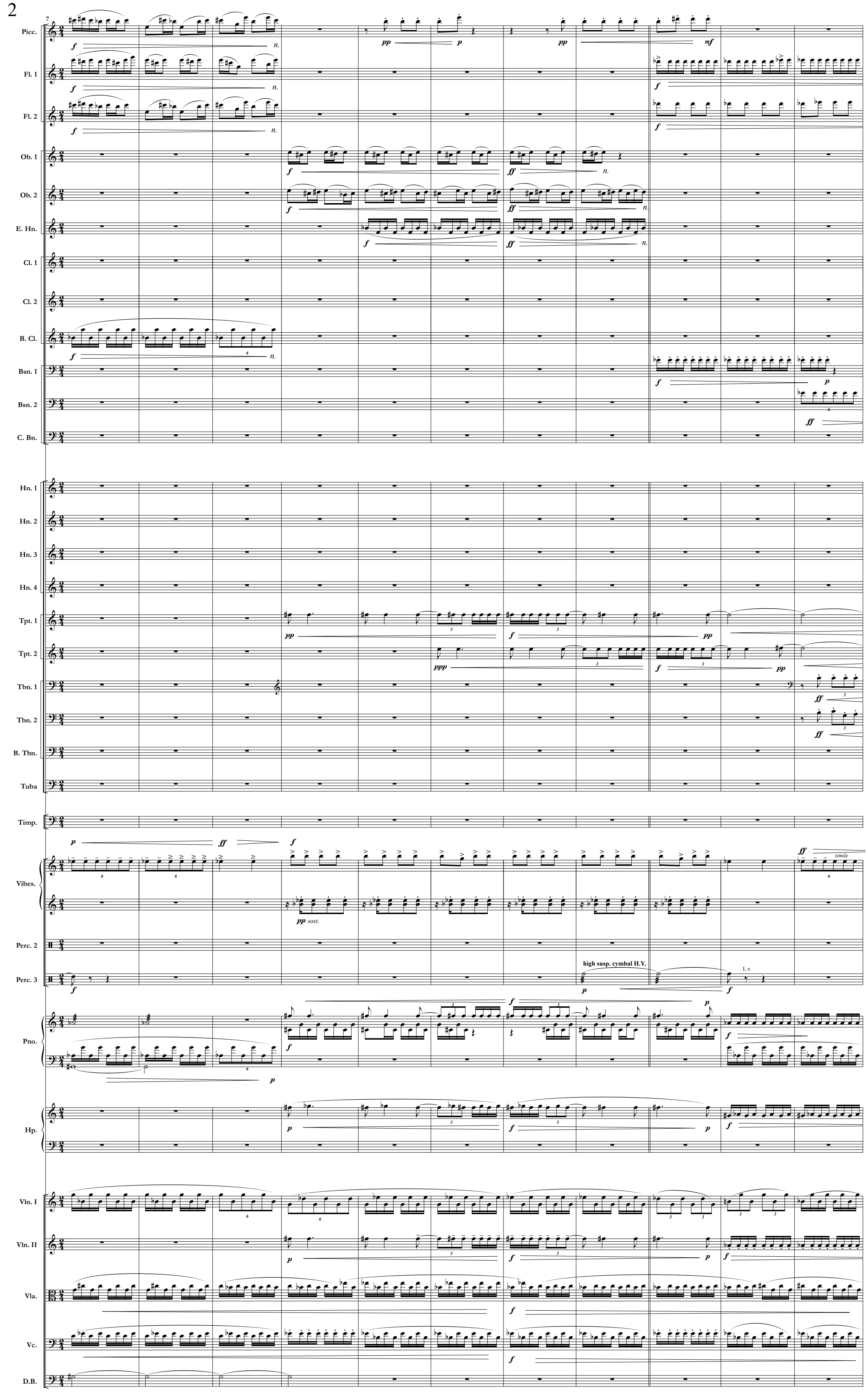
ϕ = application of a metal chains over the piano strings of indicated note (allowing for the touching the chain as well). This should be played with pedal down. This event occurs in the piano at m. 314.

▼= to the edge of the sound

SPENCER TOPEL • 917.470.7596 • 6242 Hallgarten Hall • Hanover NH • 03755

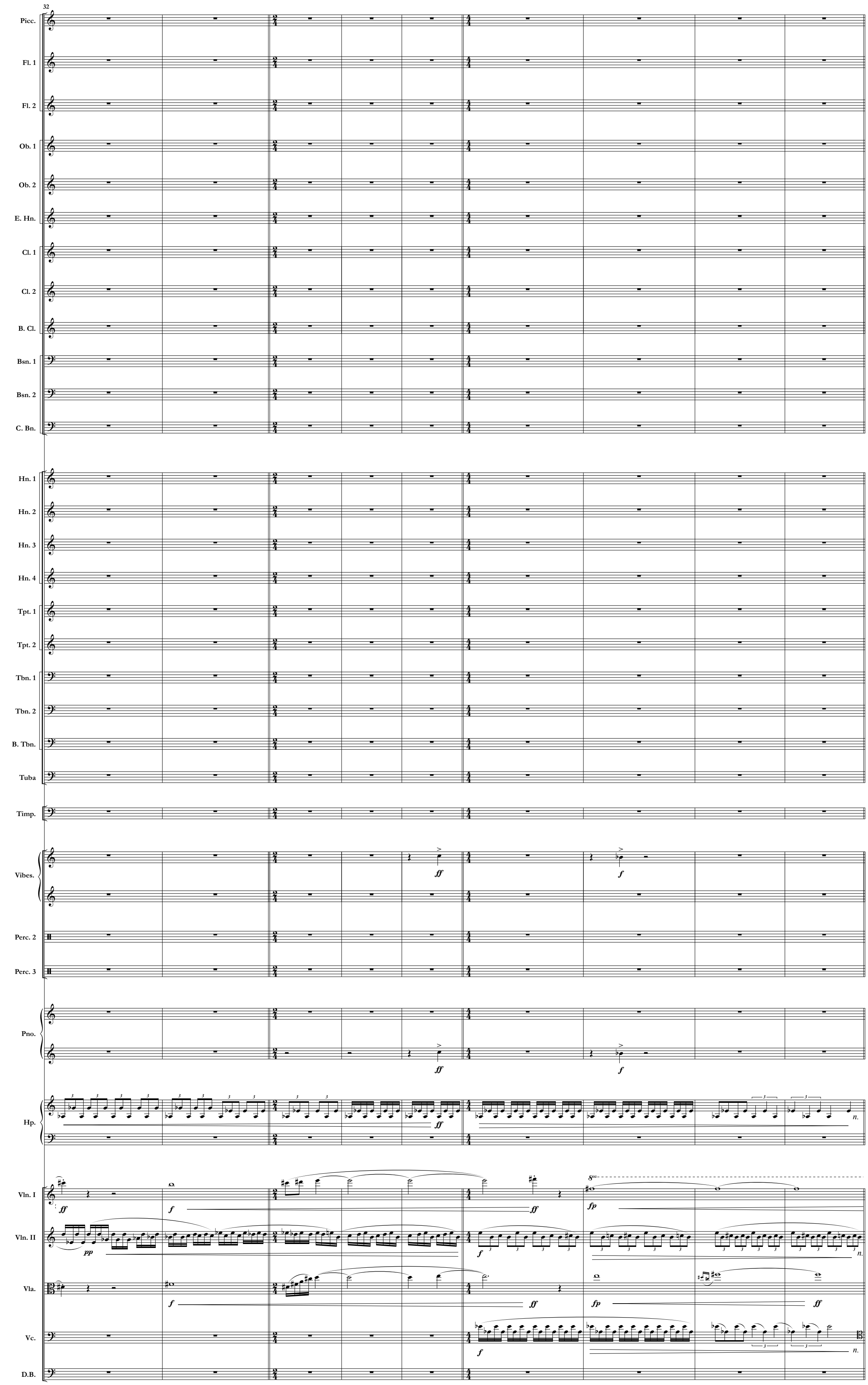
I N C E N D I O

S. Topel


Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
high susp. cymbal H.Y.
L.v.
simile
sost.
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.


18
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
crash cymbals
L.v.
bass drum
Perc. 3
Pno.
Hp.
Vln. I
sost.
Vln. II
Vla.
Vc.
D.B.


24
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
sost.
Perc. 2
Perc. 3
Pno.
Hp.
sost.
Vln. I
tutti solo
Vln. II
Vla.
tutti solo
Vc.
D.B.


32
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.

41
♩ = 84
𝅗𝅥 = 72
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
pp sost.
Tam-Tam
Susp. Cymbal (M.Y.)
l.v.
bass drum
simile


49
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
sizzle cymbal
Perc. 2
bass drum
Perc. 3
Pno.
Vibraphone
Hp.
Vln. I
Vln. II
Vla.
Vc.
pizz
D.B.


55
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
l. v.
sost.
senza sord.
tutti solo
arco


62
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
solo
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
simile
Vln. II
simile
Vla.
Vc.
D.B.


69
♩ = 84
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
tutti solo
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
sost.
Tuba
sost.
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
sost.

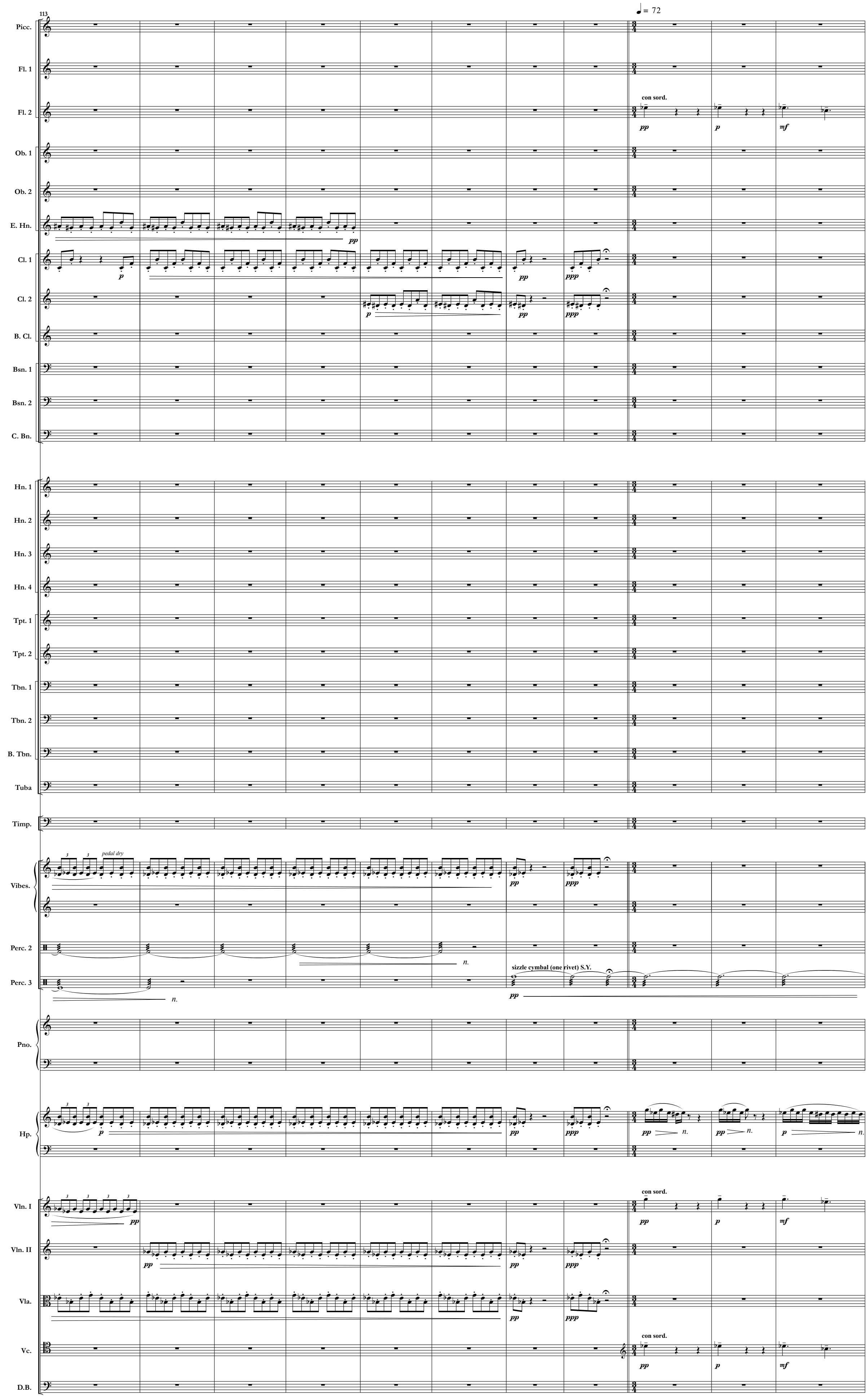

77
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
solo
molto

84
= 72
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
niente
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
niente
B. Tbn.
niente
Tuba
Timp.
Vibes.
pedal ad. lib.
Perc. 2
Susp. Cymbal (M.Y.)
Perc. 3
simile
Pno.
simile
Hp.
simile
Vln. I
div.
Vln. II
Vla.
Vc.
D.B.

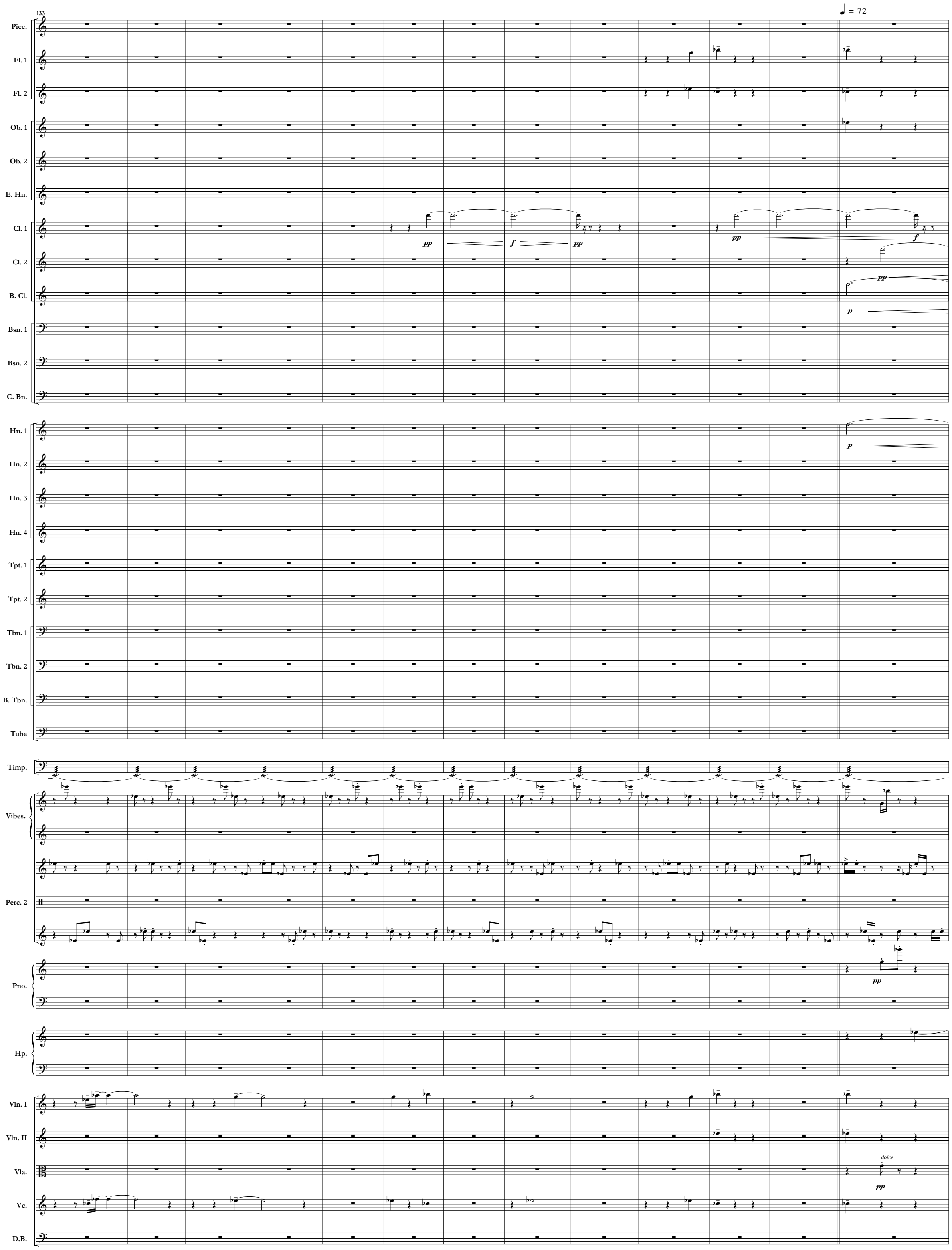

91
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
sost.
simile
Perc. 2
Perc. 3
high susp. cymbal H.Y.
L.v.
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.


99
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
crash cymbals
l.v.
Perc. 3
bass drum
Pno.
Hp.
Vln. I
sost.
Vln. II
sost.
Vla.
Vc.
D.B.


106
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
low G# to Eb
Vibes.
Perc. 2
Tam-Tam
l.v.
Perc. 3
bass drum
Pno.
Hp.
Vln. I
Vln. II
Vla.
con sord.
Vc.
D.B.


113
♩= 72
Picc.
Fl. 1
Fl. 2
con sord.
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
pedal dry
Vibes.
Perc. 2
Perc. 3
sizzle cymbal (one rivet) S.Y.
Pno.
Hp.
Vln. I
con sord.
Vln. II
Vla.
Vc.
con sord.
D.B.

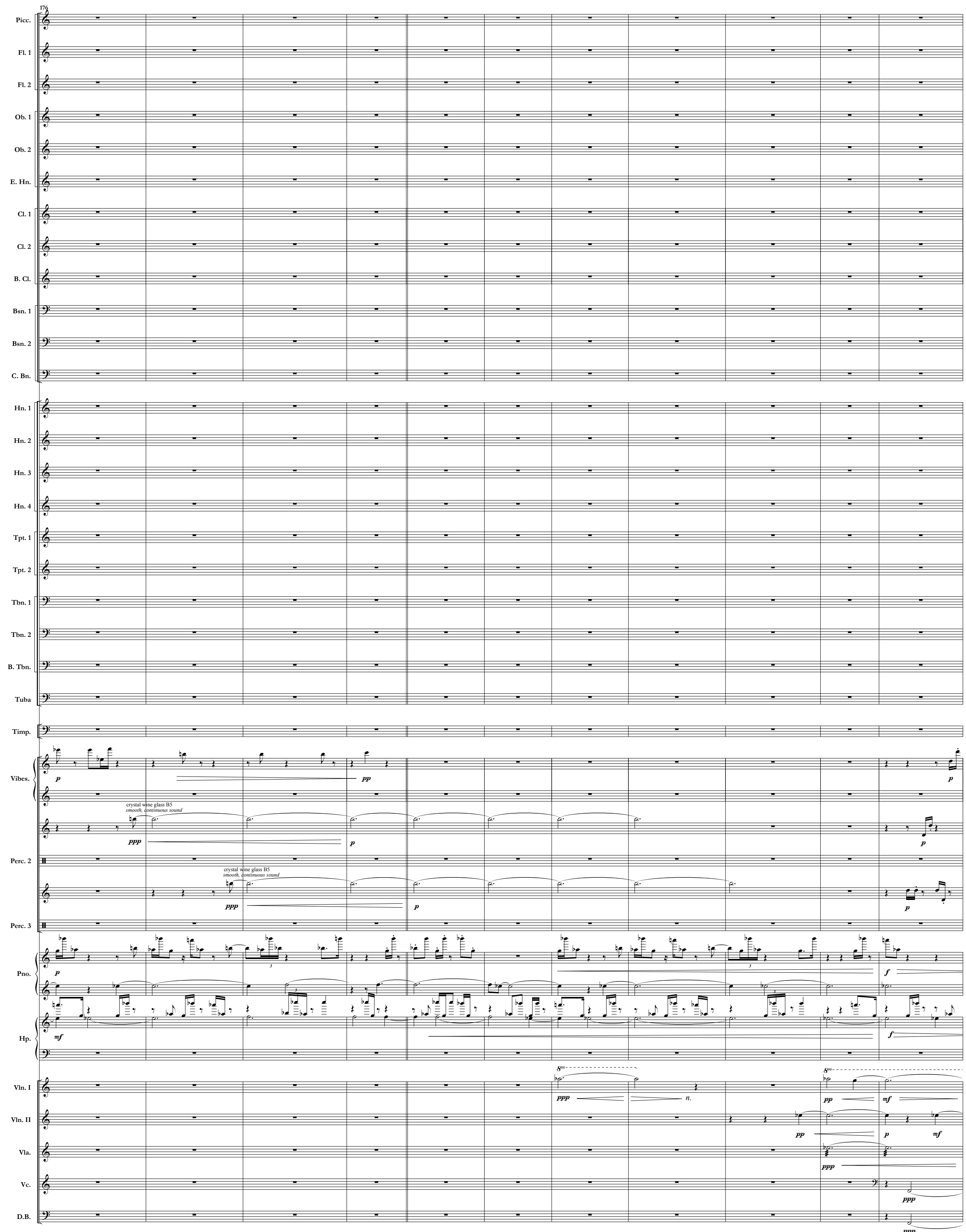

124
♩= 144
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibraphone (mallet dampen staccato notes)
p sost.
Vibes.
Glockenspiel (mallet dampen staccato notes)
pp sost.
Metal Pipes (mallet dampen staccato notes)
highest register possible
p sost.
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
col legno battuto (no pitch)
D.B.


133
♩ = 72
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
pp
f
pp
pp
f
pp
p
p
pp
dolce
pp


146
♩ = 69
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
simile
Vln. I
Vln. II
Vla.
Vc.
D.B.
senza vibrato

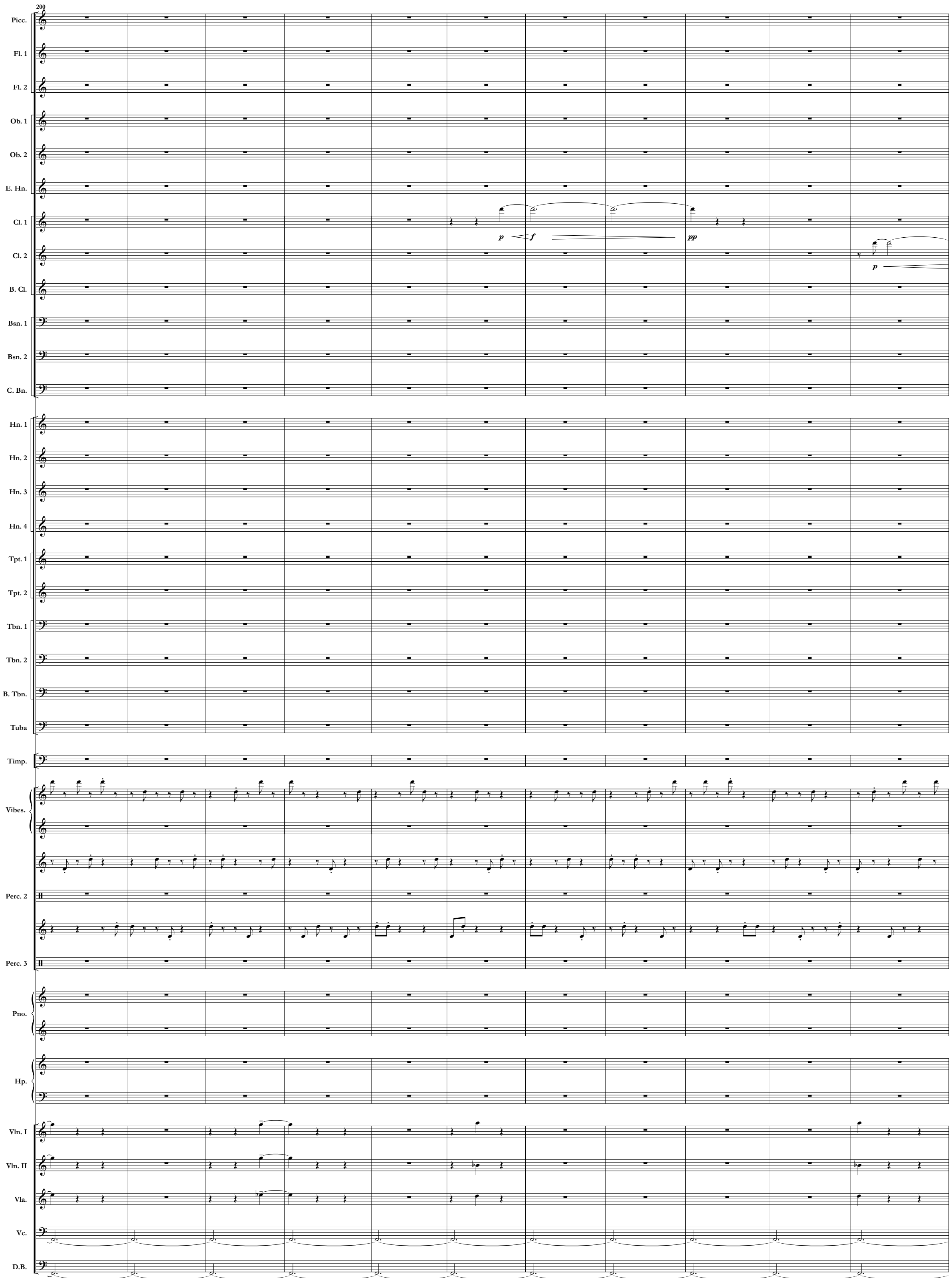

161
più accel.
♩ = 72
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
sizzle cymbal
Perc. 3
Pno.
Hp.
Come di un respiro ♩ = 76
Vln. I
Vln. II
Vla.
divisi (3)
Vc.
D.B.

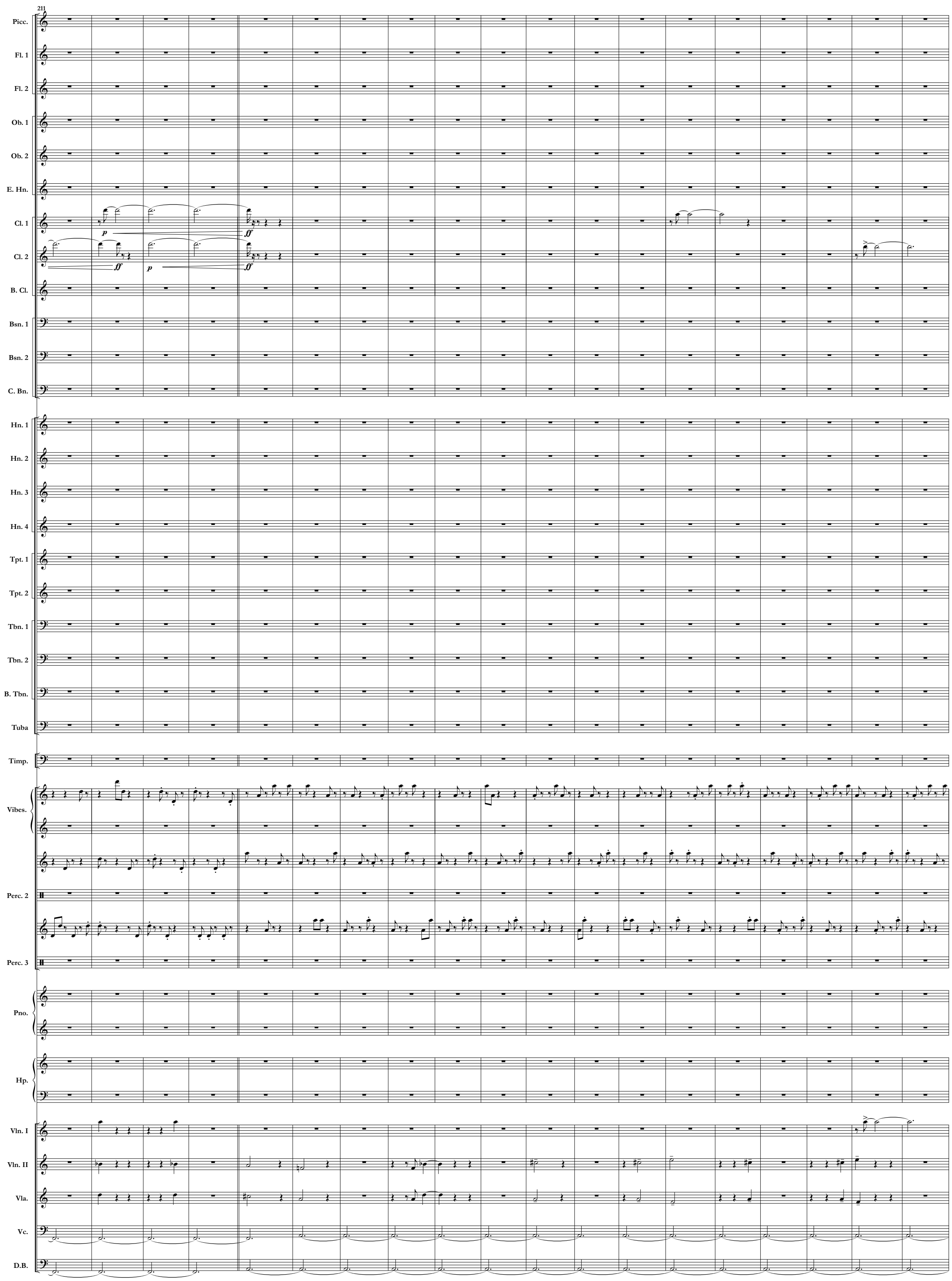

169
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
pp
fp
ff
mf
p
fff
sfz
pizz.
f

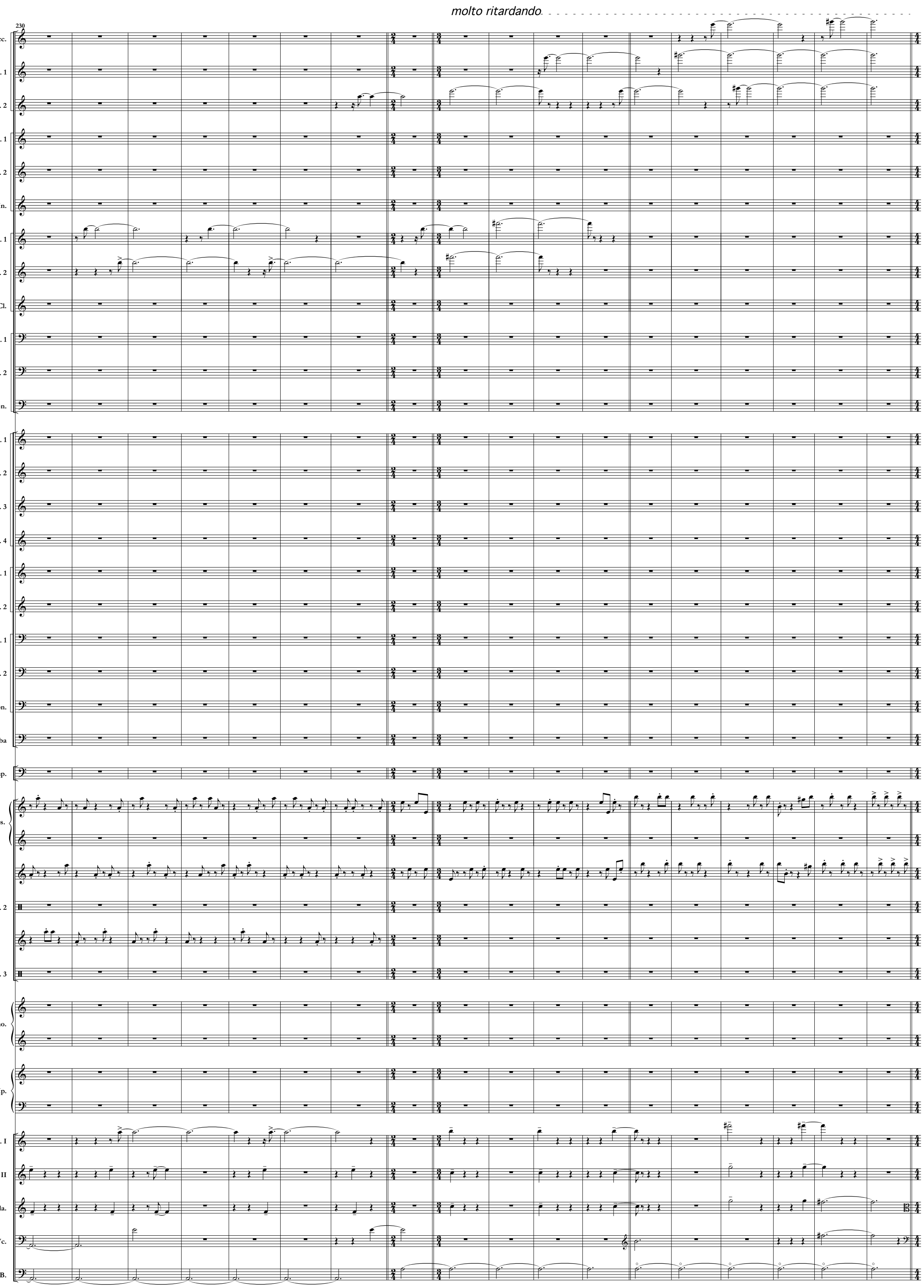

176
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
crystal wine glass B5
smooth, continuous sound
Perc. 2
crystal wine glass B5
smooth, continuous sound
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
p
pp
ppp
mf
f
n.
8va

187
♩ = 144
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Perc. 1
Perc. 2
Perc. 3
Pno.
p
pp
Hp.
p
mf
Vln. I
(8va)
Vln. II
ppp
Vla.
mf
ppp
Vc.
D.B.

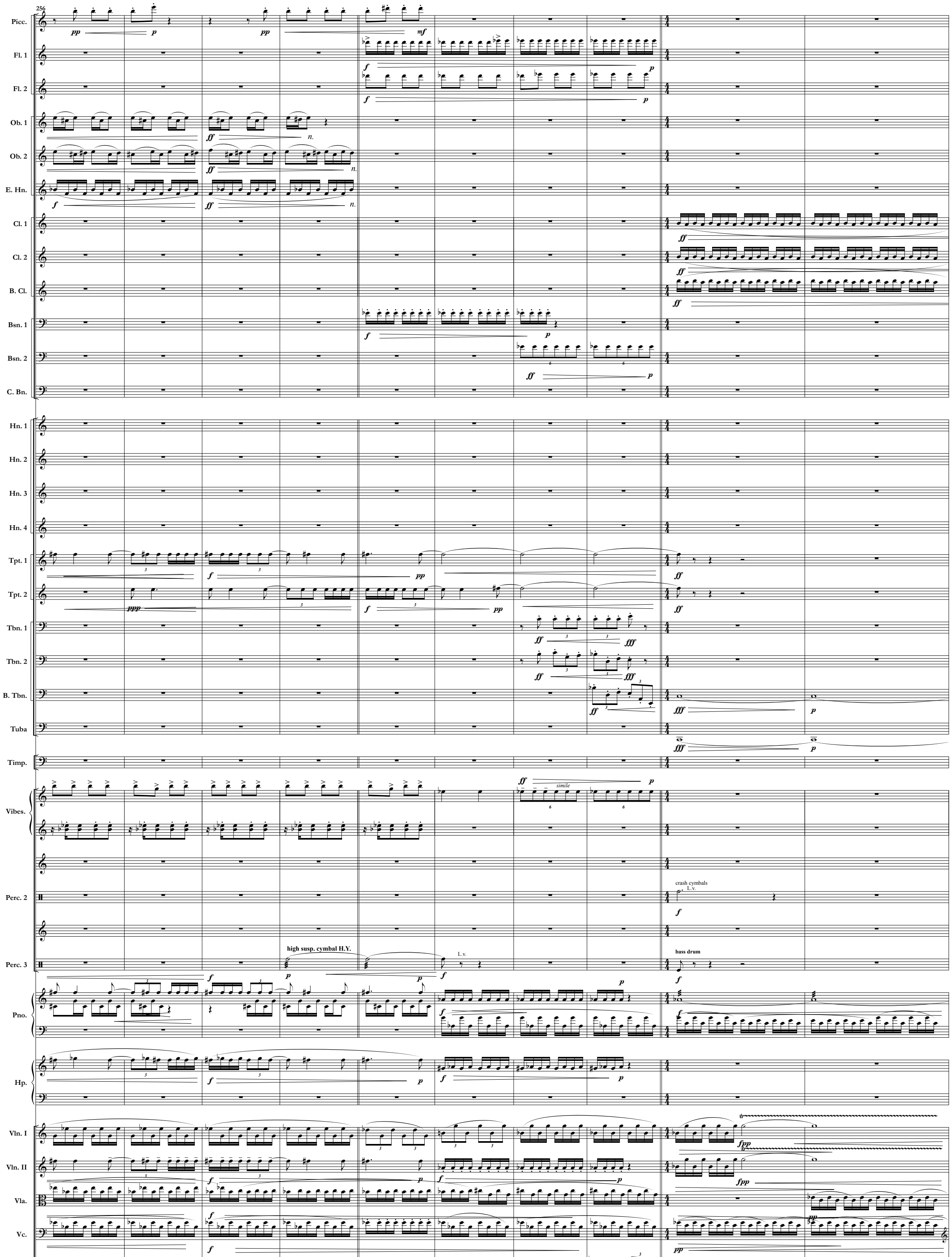

195
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.


200
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
p
f
pp
Cl. 2
p
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.

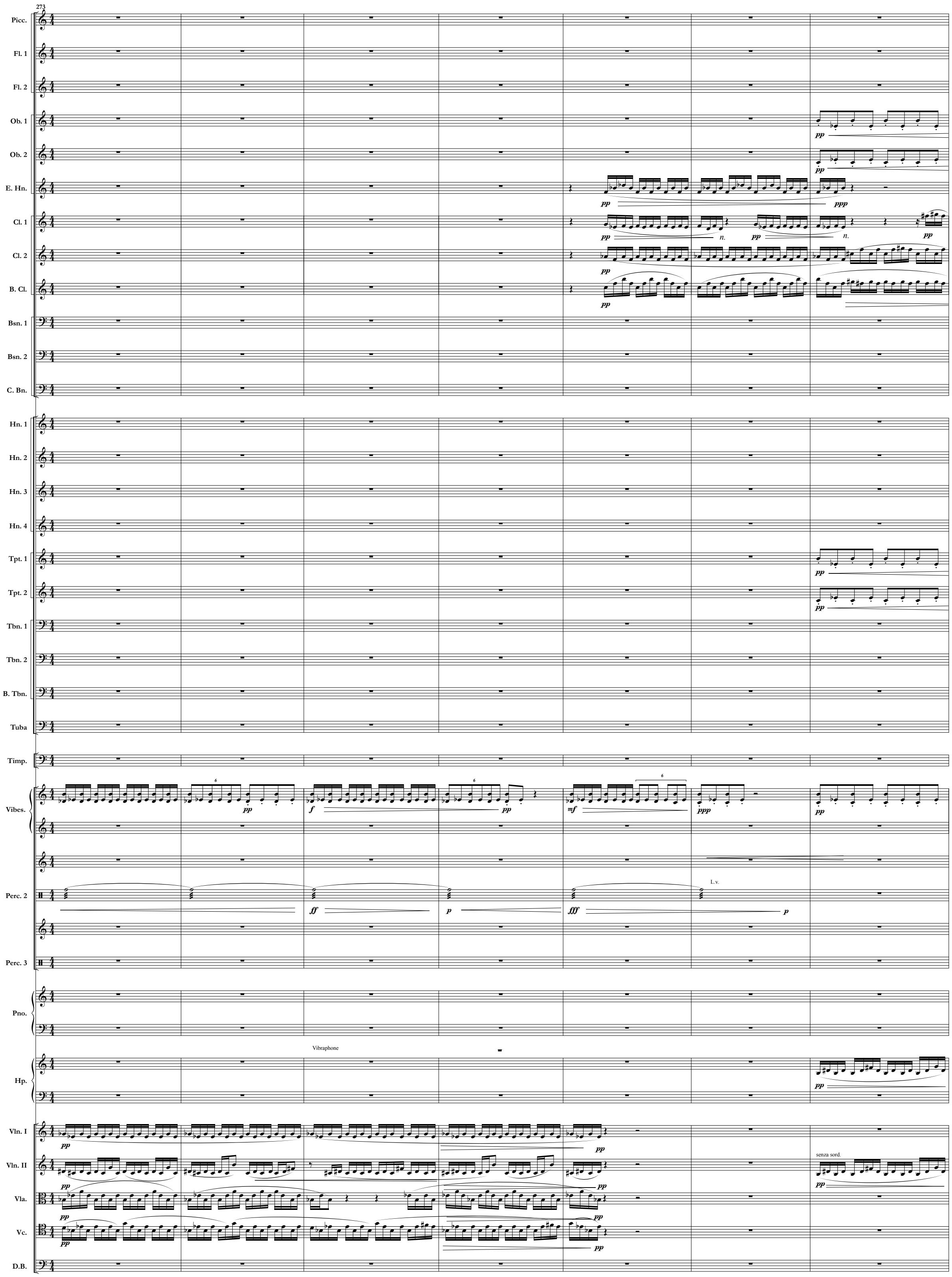

211
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
p
ff
Cl. 2
ff
p
ff
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.


molto ritardando
230
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.


♩= 144
248
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
niente
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
pedal ad. lib.
pp sost.
Perc. 2
Susp. Cymbal (M.Y.)
Perc. 3
Pno.
simile
Hp.
Vln. I
Vln. II
div.
Vla.
Vc.
D.B.


256
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
simile
Perc. 2
crash cymbals
L.v.
Perc. 3
high susp. cymbal H.Y.
bass drum
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.


266
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
L.v.
sizzle cymbal
Perc. 3
bass drum
Pno.
Hp.
Vln. I
sost.
Vln. II
sost.
Vla.
Vc.
D.B.
pizz.


273
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
l.v.
Perc. 3
Pno.
Vibraphone
Hp.
Vln. I
Vln. II
senza sord.
Vla.
Vc.
D.B.


280
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
solo
sost.
tutti solo
arco
simile


288
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
simile
Vla.
Vc.
D.B.


296
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
sost.
solo


304
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
solo


313
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
pedal dry
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
con sord.
Vc.
D.B.


322
Picc.
Fl. 1
Fl. 2
Ob. 1
Ob. 2
E. Hn.
Cl. 1
Cl. 2
B. Cl.
Bsn. 1
Bsn. 2
C. Bn.
Hn. 1
Hn. 2
Hn. 3
Hn. 4
Tpt. 1
Tpt. 2
Tbn. 1
Tbn. 2
B. Tbn.
Tuba
Timp.
Vibes.
Perc. 2
Perc. 3
Pno.
Hp.
Vln. I
Vln. II
Vla.
Vc.
D.B.
p
pp
pppp
n.
sizzle cymbal (one rivet) S.Y.

SPENCER TOPEL

CYCLICUM

for string trio and laptop (2010)

version 1, A ugust 2010

Cyclicum (2010) is a composition for string trio and live- electronics that uses feedback and comb-filters to expresses the motion of cycles and phase between different pitch fields, modeled after *Tala* in indian classical music. Excitation of inharmonic parametric filters tuned to different frequency patterns influences coloration and phase cancellation in higher frequencies. Sound spatialization is employed to control feedback and distribute the auditory scene of the integrated experience as well as contribute to the choreography of cyclical motion of auditory partials.

Technical Note

<u>Software</u>
OS X 10.5.8 or later running Max 5
Princeton Univ. percolate externals or author supplied recompiled version
ICST Zurich Ambisonics Suite
Cyclicum max patch

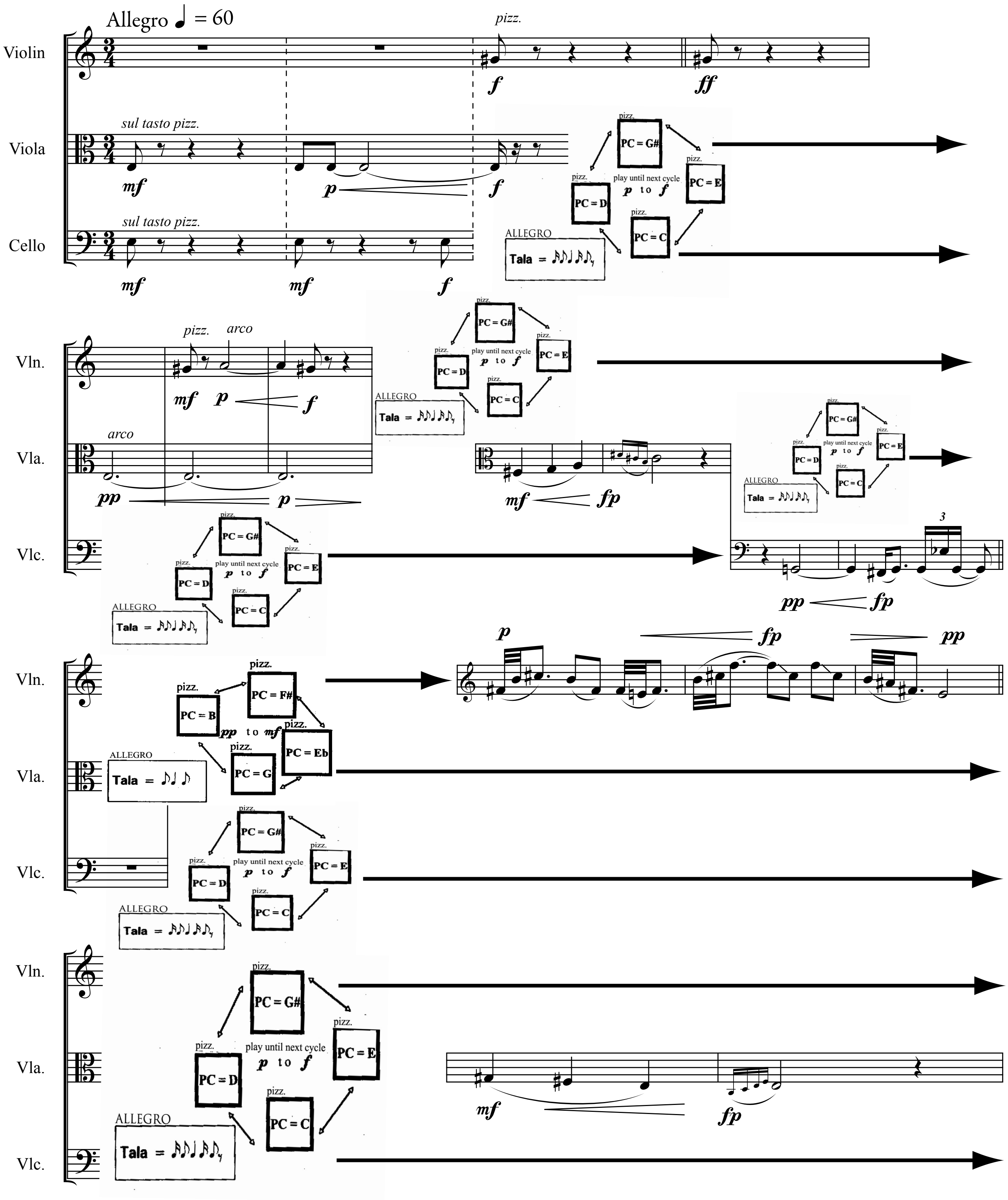
<u>Hardware</u>
2008 Intel Mac or better
Three Danish Pro-Audio mini-microphones (ideal)
digital interface (DA/AD) to the computer multi-channel speaker distribution array

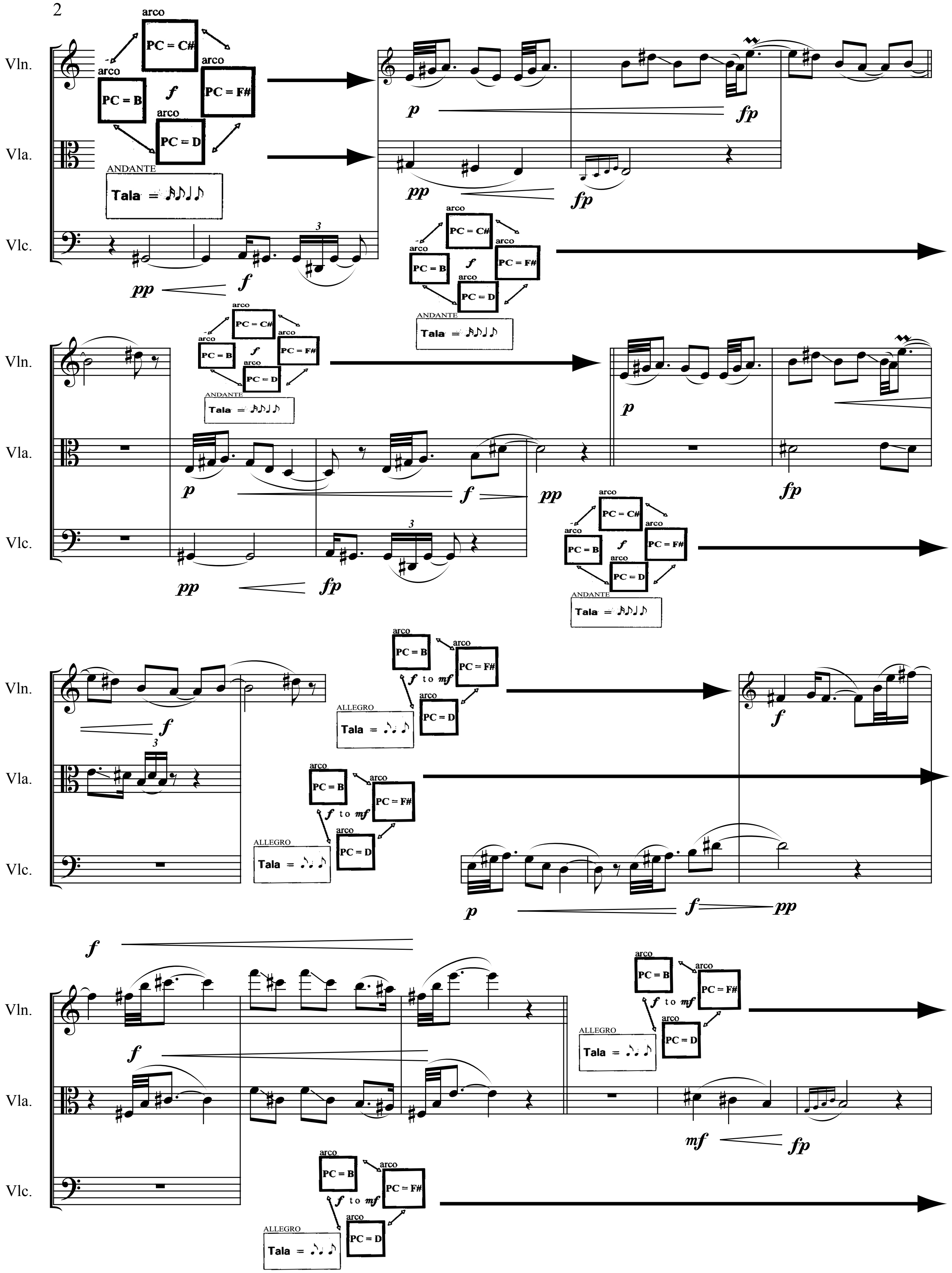
<u>Humanware</u>
string trio (violin, viola, and cello)
Laptop Operator (recommended)

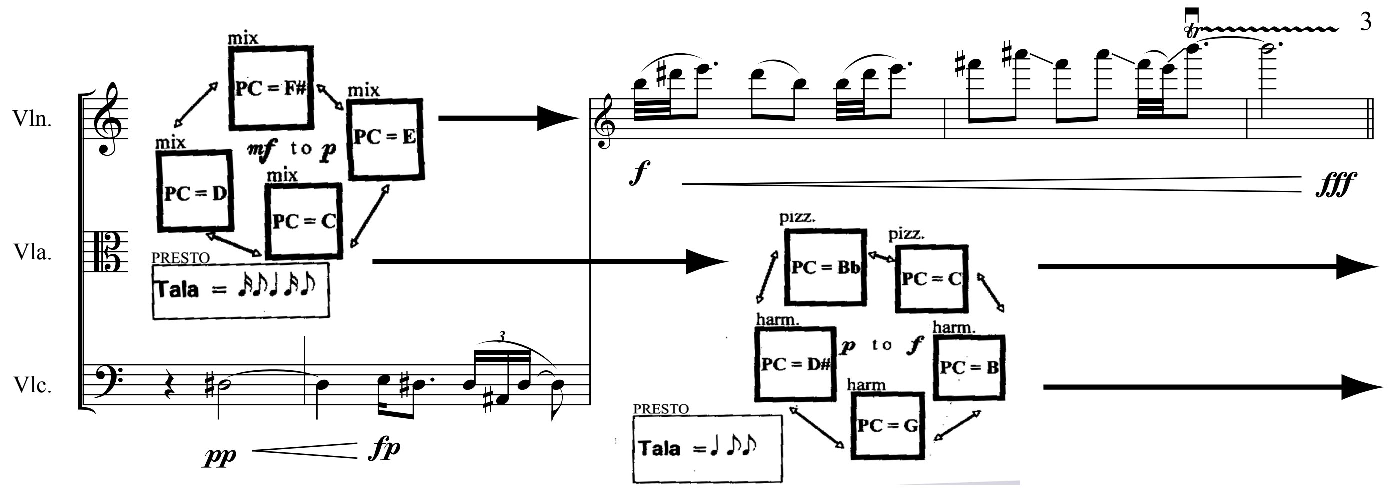
Cyclicum (2010) Performance Notes

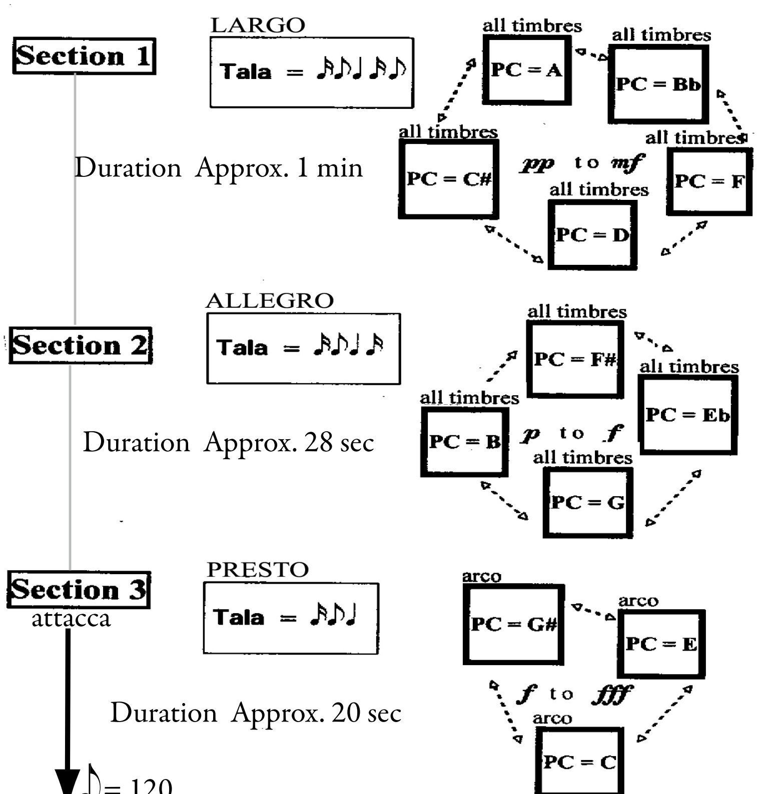
- Cycles consist of boxes, arrows, talas, and indications.
- A player may start at any box and move from box to box only if it is connected by an arrow.
- Talas indicate both the relative tempo (e.g. Allegro, Largo, etc.) and a repeating rhythmic pattern, where the notes change while the rhythm stays relatively the same. Flexibility can be taken in executing the pattern, but variants should remain somewhat similar to the original.
- Indications appear above the box that instruct what kind of articulation of the note should be used.
- When the indication mix or all timbres appears, a player is free to move between different styles of playing their instrument.
- Notated music appears in all three parts. When these fragments occur, the player must make a smooth transition between the cycles and the fragments.
- sound files played back during performance are either recorded in realtime or pre-recorded. They can be operated by an additional performer on laptop or setup to play back automatically.
- Written sections of music (the beginning and end) plus the fragments should have a constant tempo of quarter-note equals sixty.

CYCLICUM


Allegro ♩ = 60
Violin
Viola
Cello
pizz.
sul tasto pizz.
mf
p
f
ff
pizz.
PC = G#
PC = E
PC = D
PC = C
play until next cycle
p to f
ALLEGRO
Tala =
Vln.
Vla.
Vlc.
arco
pp
fp
PC = F#
PC = B
PC = Eb
PC = G
pp to mf
3


Vln.
Vla.
Vlc.
arco
PC = C#
PC = B
PC = F#
PC = D
f
ANDANTE
Tala =
p
fp
pp
3
f to mf
ALLEGRO
mf





Description:

This section consists of the same improvisation circles as described before with a few exceptions:

1) The dashed lines indicate that a player can occupy the same pitch class as another player.

2) In general this section should emphasize buildup to the unison melody on the next page.

3) As the reduction of pitch classes occurs, greater empasis should be placed on the tala structures as to create a sense of drive and intesity.

4) between section 2 and 3, players should be arco and by section 3 playing marcato and/or near the frog

5) all timbres referes to using any possible color on the instrument to produce the indicated pitch class.


4
Vln.
Vla.
Vlc.

SPENCER TOPEL

Aqua Regia

for flute, bass clarinet, violin, viola, cello, piano, and percussion
2008

version 1, July, 2008

Transposed Score

Aqua Regia

dur. 6 min.

S. Topel


B
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
sost.
arco
C
pizz.
arco


24
D
Fl.
B. Cl.
Crt.
glockenspiel
Mrmba.
Pno.
Vln.
pizz.
Vla.
pizz.
Vc.
pizz.
30
E
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
arco
con sord.
Vla.
arco
con sord.
Vc.
arco
con sord.

37
F
Fl.
B. Cl.
crotales
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
46
G
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.


H
K
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
senza sord.
senza vib.


63
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
ord.
molto legato
70
L
M


81
N
P
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
89
Fl.
B. Cl.
Mrmba.
Pno.
Vln.
Vla.
Vc.
pizz.


95
Q
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
arco
100


105
R
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
110
S
sus. cymbal med.
crotales


116
T
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
127
U
= ca. 72
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.


V
poco á poco rit...
solo whistling
molto rit.
W
= ca. 96
133
139
X
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
p
pp
f
mfz


Y
144
Fl.
B. Cl.
Crt.
Mrmba.
Pno.
Vln.
Vla.
Vc.
150
pizz.
senza vib.
senza vib.

SPENCER TOPEL

Elementary Sources

for string quartet and laptop (2010)

MVT. 1

version 2, September 2011

Elementary Sources (2011)

Program Note

A central concept for this work was whether it was possible to acoustically re-create a simple audio source recording with an entirely different set of sources such as a string quartet. This is not unlike the concept Gérard Grisey presents for his landmark work “Partiels” (1975), where he describes the concept of using the orchestra for the purpose of Macro-synthesis, where each instrument of the orchestra contributes to a larger sound object, rather than distinctive sections and instrumental colors.

The audio source material is a recording created by a local artist of bell-halves she constructed from brass hitting concrete an specific time intervals. The decomposition of these sources results in consistent timbral components relating to the various distinctive parts of the original audio. Namely, the cement “clicks”, resulting from the moment the metal hit the ground, the in-harmonic partials from the metal as it rang, and the oscillation, or “wobbling”, of the halves as they came to rest on the cement. Through transformation of components, and blending of the acoustic string quartet with a combination of sample playback and live-electronics, the piece navigates the arthroscopic details of the original sounds while striving to find expression between their respective spaces.

Technical Note

Performers: 2 Violins, Viola, and Cello (string quartet)
Laptop: Running Ableton LIVE Suite 8.0 or higher
Microphones: Danish Pro-Audio Mini-capsule Microphones 4x
Speaker Arrangement: Stereo output, preferably behind or near the performers to double as monitoring.

Score

Elementary Sources

brass bell-halves

♩ = 72

S. Topel

c.a. 6 min., 30 sec.


12
Vln. I
Vln. II
Vla.
Vc.
sul pont. pizz.
mf
sfp
f
sul pont. pizz.
mf
18
p
Glissando
Gliss
Glissando
Glissando
gliss.
gliss.


24
8va
Vln. I
ppp
f
pp
f
Vln. II
ppp
f
pp
f
Vla.
ppp
f
pp
f
Vc.
pp
f
ppp
f
28
8va
Vln. I
p
ff
p
f
pp
ff
Vln. II
p
ff
p
f
pp
ff
Vla.
p
ff
p
f
pp
ff
Vc.
p
ff
p
f
pp
ff


32
Vln. I
Vln. II
Vla.
Vc.
sfz
sul pont. pizz.
35
pp
p
fp
f
Gliss


39
Vln. I
gliss.
gliss.
gliss.
fpp
Vln. II
gliss.
gliss.
gliss.
gliss.
p
fpp
Vla.
gliss.
fp
ff
Vc.
glissando
p
fp
ff
44
Vln. I
ff
n.
Vln. II
ff
n.
Vla.
sffp
n.
Vc.
sffp
n.


48
Vln. I
Vln. II
Vla.
Vc.
mf sost.
sul pont. pizz.
mfz
mf sost.
sul pont. pizz.
mfz
f
mf
53
Vln. I
Vln. II
Vla.
Vc.
mf
f
n.
arco
senza vib.
pp
p sost.
arco
senza vib.
f
p
f
n.
p sost.
pizz.
strimpellare
5
p


59
Vln. I
Vln. II
Vla.
Vc.
simile
5
mf
5
7
64
Vln. I
Vln. II
Vla.
Vc.
simile
5
7
f
7


69
Vln. I
Vln. II
Vla.
Vc.
arco
75
Glissando
p sost.
pizz.
arco
Glissando


79
Vln. I
Vln. II
Vla.
Vc.
fp
n.
83
cantabile
p
mf
pp
f
mf sost.
Gliss


88
Vln. I
Vln. II
Vla.
Vc.
gliss.
gliss.
gliss.
glissando
gliss.
f
fff
93
G.P.
n.
pp


98
Vln. I
Vln. II
Vla.
Vc.
pp
f
p
f
n.
mf
104
sffz
sul pont. pizz.

Three Preludes

by Spencer Topel

2007

Three Preludes 2007

Performance Instructions and Notes

Prelude I (3 min.)

- Open beam notation allows the performers to focus on the changes of notes. This translates into incremental glissandi in the violin (for each repetition of the bow stroke) and conventional glissandi for the cello, from each note head to the next. Avoid over-articulation in the violin part.
- Shaded dynamics are meant to express generalized intensity. For the cello, this means that dynamics, vibrato, and sounding point are to be expressed together. Likewise, the violin expresses dynamics, bow-stroke, and sounding point together.
- Closed non-shaded dynamic indication that happens only in the second system indicates Rubato towards the climax.
- Two distinctive bow-strokes appear in the first prelude. The first being a conventional bow crossing method that is indicated by a circled number 1. A second, more prevalent, bow-stroke is also indicated by a circled number 2. This stroke is widely regarded by violinists as the J.S. Bach E Major Partita (BWV 1006) bowing that appears in the Preludio movement.
- In the violin part, the open "E" string should sound throughout.
- Care should be taken in synchronizing big arrival points. One of the two players should assume the responsibility of cueing each entrance.
- Notes with fermatas indicate rest points. Discretion should be taken as to how long each fermata should be, but good musical judgment will suffice.
- Non-filled notes indicate emphasis.

Prelude II (2 min. 10 sec.)

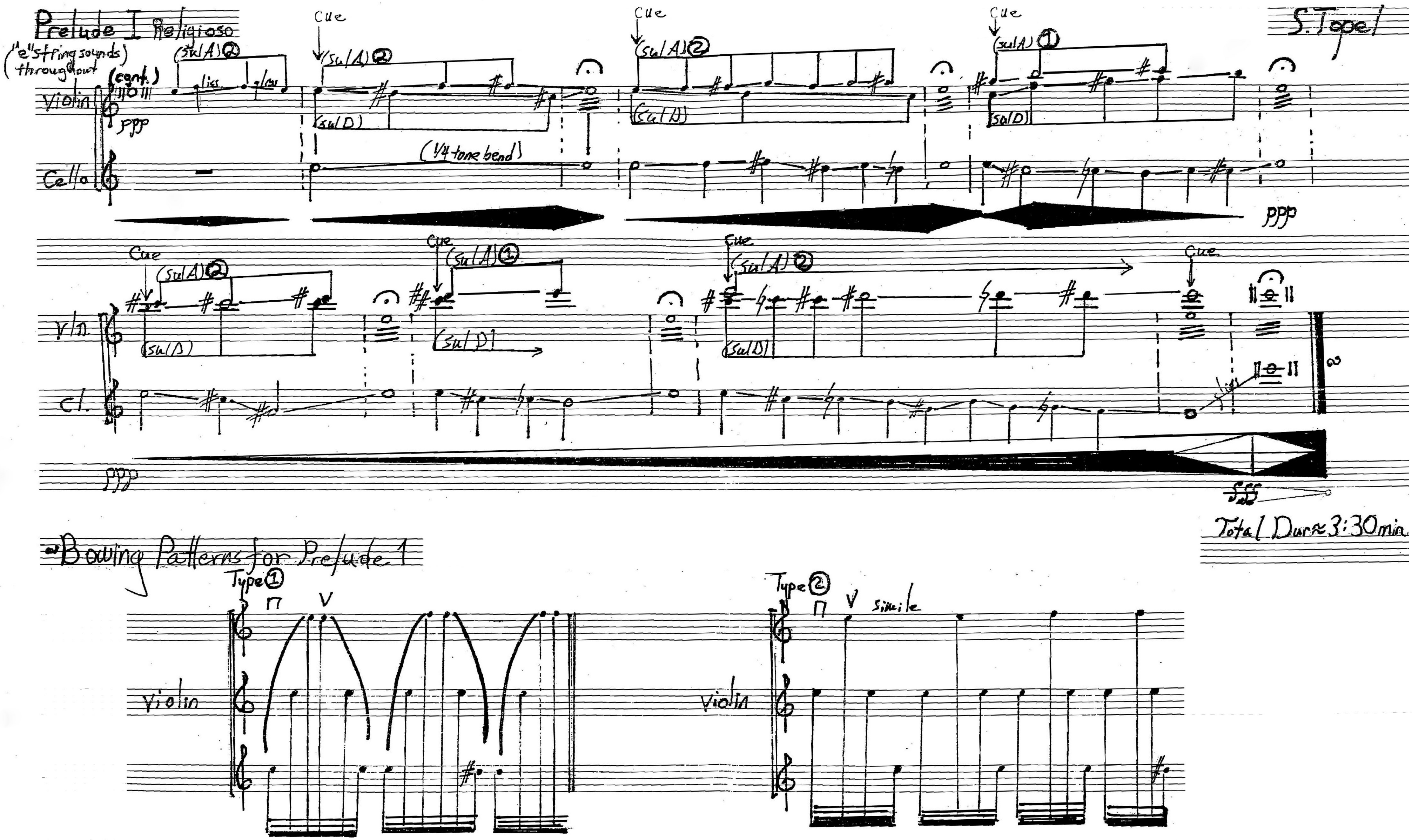
- Note values apply in this prelude, unlike the first prelude.
- Additional staves above and below the violin and cello part indicate the left hand pizzicati.
- Pizzicati at the end of the prelude should fade away, at the player's discretion.
- In the m.5, sul pont. should gradually become sul tasto in each subsequent chord.

Prelude III (3 min. 30 sec.)

- Notes followed by an arrow indicate that a figure should keep repeating for either as long as indicated (with an circled number to indicate repetitions) or by duration. The major exception to this is in the last system, which should be based on the arrival time of the synchronization of the two lines followed by a relatively equal proportion of music following the synchronization. The cello initiates the arrival of the last measure of music.
- The violin figures written as thirty-second notes should be played as fast as possible.
- In system three, the bracketed music must be played together. The grace-notes in the violin part should be added in relation to the proportions of the system. The total duration of this phrase should be approx. eight seconds.
- m. 12 contains a shorthand notation for the cello. The first four notes sets the pattern and each note head following should be played by first adding an "E" sixteenth followed by the note indicated.
- Dashed notes in the violin part in mm.12-18 should be played simultaneously with the figure, but not held. Rather, the notes should be touched upon, similar to the approach to the technique a sitar player uses to keep adjacent tones resonating by lightly touching the note in a non-rhythmic fashion.
- The fermata with an infinity symbol reminds the performers that this fermata must be held a VERY long time and that the tremolo, grace note trill figure in the violin, and the general intensity should continue to hold until the players can hold it no longer.

Program Note

The First performance was at the Palazzo Chigi in Sienna Italy by Duccio and Vittorio Cennanti in the summer of 2007. These pieces explore the relationship between the timbres of the violin and cello and are an acoustic representation of aesthetics developed in the electronic studio. Both of these instruments hold great personal significance to me in the playing and writing of chamber music. This work is dedicated to composer Steven Stucky. –*Note by the composer*

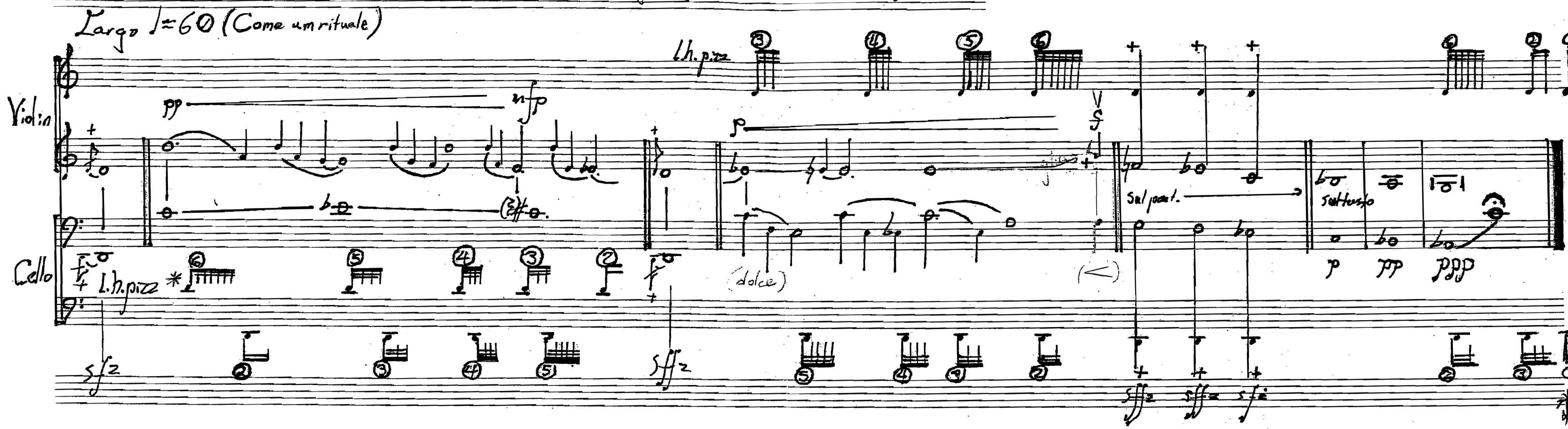


Prelude I Religioso
S. Topel
("e" string sounds) throughout
(cont.)
Violin
Cello
ppp
(sul A) ②
gliss
Cue
(sul D)
(1/4 tone bend)
(sul A) ①
Vln.
Cl.
Total Dur ≈ 3:30 min
Bowing Patterns for Prelude 1
Type ①
violin
Type ②
simile

Prelude II

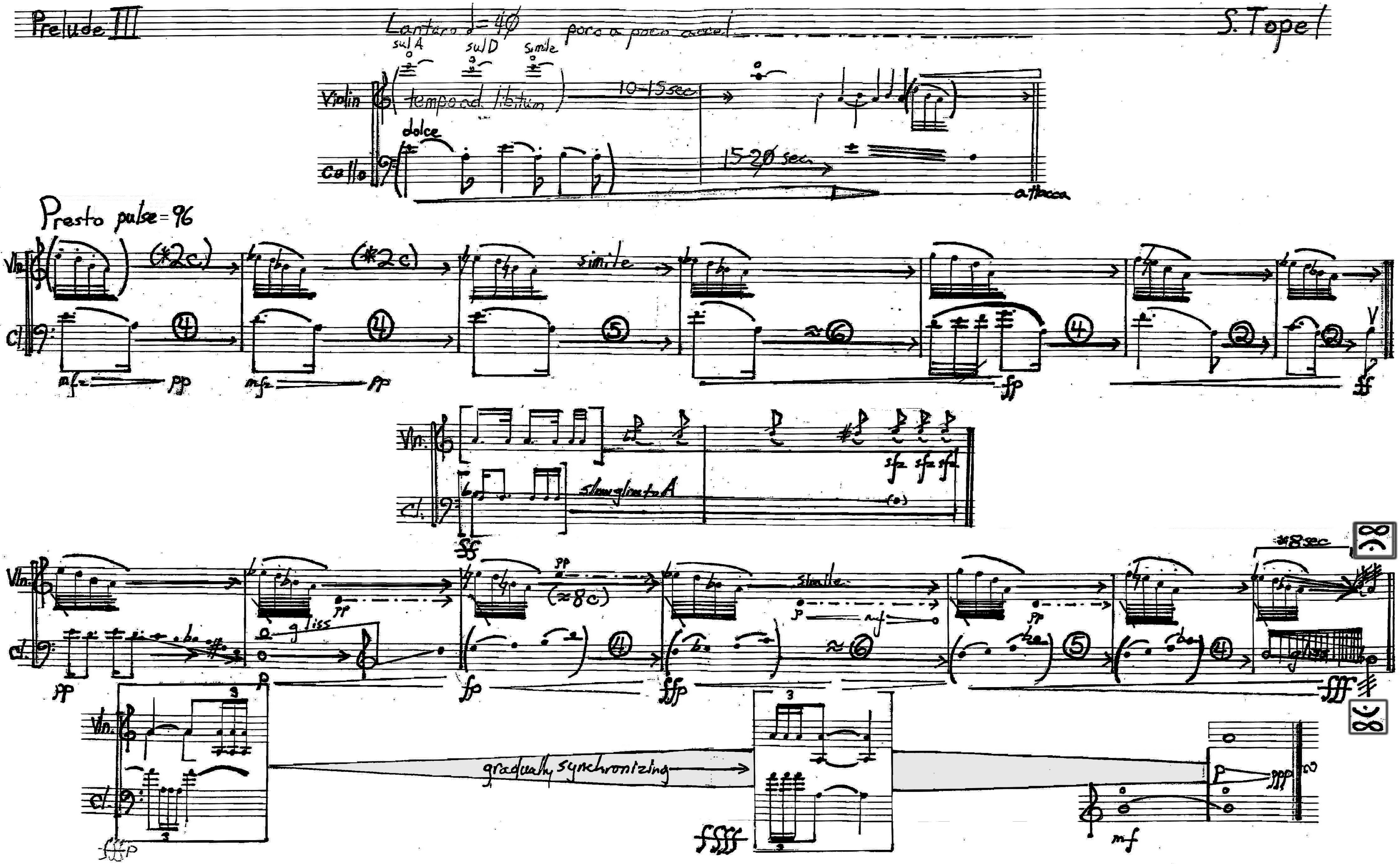
S. Topel

scordatura cello C → D (IV) or an octave higher with violin performing l.h. pizz on top voice of cello.



* if this is not possible, then either the violinist can play this or it can be omitted.

** ad libitum, slowly fade away repeating pizz.


Prelude III
Lantano = 40
poco a poco accel.
S. Topel
sul A
sul D
simile
Violin
tempo ad libitum
10–15 sec
dolce
Cello
15–20 sec
attacca
Presto pulse = 96
Vln.
Cl.
(≈2c)
simile
mf
pp
fp
ff
sf
slow gliss to A
gliss
(≈8c)
≈8 sec
p
fffp
fffp
fff
gradually synchronizing
ffff
mf
ppp

EXTENSIBLE SPECTRALISM: REVEALING LATENT STRUCTURES IN MUSIC AUDIO FOR COMPOSITION, ANALYSIS, AND RETRIEVAL

Part II

A Dissertation
Presented to the Faculty of the Graduate School
of Cornell University
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Musical Arts

by
Spencer Topel

May 2012

EXTENSIBLE SPECTRALISM: REVEALING LATENT STRUCTURES IN MUSIC AUDIO FOR COMPOSITION, ANALYSIS, AND RETRIEVAL

Spencer Topel, D.M.A.

Cornell University 2012

Music exemplifies the repetitive patterns in nature. These patterns lend a distinctiveness to sound sources that make them identifiable. In audio analysis, this information can be accessed by using a process called spectral decomposition. This dissertation evaluates the ways in which spectral decomposition techniques can yield a different way to understand music.

Using an interdisciplinary framework incorporating Albert Bregman's *Auditory Scene Analysis* with traditional music and computational analysis methods, spectral decomposition techniques are employed in the following four activities: a digital-musicology study of Grisey's *Partiels*, and John Cage's *Sonatas and Interludes*; a timbre-rhythm groove retrieval analysis on a new dataset named ISHKUR, a discussion of a repertoire of music written using spectral decomposition techniques, and future directions for music research and aesthetics.

The main conclusions drawn from this research illustrate the versatility of latent structure analysis to activities beyond source separation and argues for a perceptual foundation for a post-spectralist approach to music.

BIOGRAPHICAL SKETCH

Spencer Topel's research covers range of activities relating to digital arts and humanities, including interactive and live-electronic music, installation and sound-art, digital musicology, music information retrieval, and source separation.

His music has appeared on concert programs in venues such as Issue Project Room, Brooklyn NY, Orchestra Hall, Minnesota, the Chiesa di Sana Caterina Treviso in Venice, Italy, the 2008 Aspen Music Festival, Chigiana Festival in Siena Italy in 2007, at Alice Tully and Weill Concert Halls in New York, and in Tokyo City Opera Hall. In May 2008, his music was featured on a concert tour of Turkey sponsored by the Turkish Cultural Center NY, featuring twelve musicians from around world. In 2011-12, Mr. Topel's music appears on programs including MATA, the International Society of Music Information Retrieval, Sound and Music Computing, Bridge the Gap Chamber Players, and the Celebration of Music at Dartmouth. In July, Spencer Topel will be appointed Assistant Professor at Dartmouth College.

ACKNOWLEDGEMENTS

I owe a deep and heartfelt thank you to my dissertation committee, consisting of Dr. Steven Stucky, Dr. Michael Casey, Dr. Roberto Sierra, and Dr. Kevin Ernste. A special thanks to Dr. Stucky, for his kindnesses and willingness to approve this ambitious idea for my dissertation project, and to Dr. Casey who guided me through the research and writing process, which was once an unrealistic dream of creating an interdisciplinary approach to new analysis and composition techniques in music. Deepest thanks also to Dr. Ernste, for his unyielding optimism and beyond-his-years mentorship, and to Dr. Sierra for his encouragement and rich insights into composing. I also owe thanks to all of the amazing conductors and musicians who have performed my music over the years, including the Minnesota Orchestra and maestro Osmo Vänskä; Vittorio and Duccio Ceccanti for their near-perfect performance of the Three Preludes for Violin and Cello; to Fredric Lacroix, for inspiring me to continue writing for piano; and to my most recent inspiration, Voxare Quartet. I also want to thank Graeme Bailey for his gracious generosity and his friendship; to Peter Brinkmann for our weekly meetings in 2009, which proved enormously valuable in developing my thinking, and to David Plans Casal, who is one of the most creative people I know; and to my past mentors: Christopher Rouse, Azio Corghi, Aaron Kernis, Tomas Svoboda, Samuel Adler, and Anders Hillborg, for encouraging me to "keep being curious!" Finally, There are no words to express the gratitude and appreciation I have for my wife, Maria Endrinal, who supported me through this entire process, and to my son William, who witnessed the end of it. I am deeply grateful to my family, my parents, Robert and Jeannette, and my brother, Ross, who listened to what was perhaps an infuriating number of preliminary versions of the ideas presented.

for Maria and William.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivations

Music exemplifies the repetitive patterns in nature (Batteux 1760). These patterns lend a distinctiveness to sound sources that make them identifiable. In audio analysis, this information can be accessed using a process called spectral decomposition. In music and audio signal processing research, spectral decomposition is commonly associated with source separation, in which an algorithm recovers the original sources from a mixture (i.e., extracting each instrument from a recording of an ensemble or band) (Barry et al. 2005, Cardoso 1989, Plumbley et al. 2007, Virtanen 2007).

Instead of searching for original sources in a mixture, this dissertation investigates the application of spectral decomposition techniques to music composition, rhythm pattern similarity analysis, and spectralist techniques to discover if this type of analysis can yield a different, new way to understand music. There are two broad motivations for this work: (a) to devise a musicology framework using spectral decomposition analysis and (b) to explore a concept called latent structure in music audio. Latent structure is defined as the relationship and arrangement of spectral patterns contributing to the formulation of conspicuous elements in an auditory scene that might otherwise be hidden by primitive auditory processing.[1]

[1] In auditory scene analysis (ASA), Bregman (1994) described listening as two cognitive and parallel processes: (a) conscious and (b) unconscious listening.

In other words, this type of analysis may enable us to explore aspects of music that might otherwise be subsumed in listeners' conscious experience of music. The results may be musically intuitive (e.g., the extraction of components that sound dissimilar or an entirely new interpretation of a musical idea).

1.1.1 Application of Spectral Decomposition in Music Analysis

Spectral decomposition can be described as the process by which a magnitude spectrogram separates into subspaces (Casey and Westner 2000), or mixing matrices (Hoffman et al. 2010). In tasks where sounds are not treated as mixtures, significant information for music creation can be determined and manipulated directly from a time-frequency representation, such as a spectrogram, by using software such as OpenMusic (Assayag et al. 1999) and Spear (Klingbeil 2005).[2]

In contrast to prior spectral techniques, the analysis methods described in this dissertation may be used to access independent features that are not as well understood in the context of music analysis and composition. One emergent question is how useful or reliable are spectral decomposed component representations in music analysis. If using decomposed spectral components improves current analysis tasks–both general and specific–then presumably they can be applied to less discriminative tasks, such as composition and improvisation.

This problem is addressed in relation to creativity, by looking at the manipulation and transformation of resynthesized spectral decomposition components, and research, where rhythm pattern similarity is examined in the context of tim-

[2]It is important to note that in a Schafferian (1967) framework nearly any acoustic source can be considered a mixture because it is comprised of discrete sounds that Schaeffer identified as Objets Sonore.

bre (Qingyuan et al. 2011). As explained in chapter 4, early results suggest that content-based groove retrieval is better than canonical sub-band decomposition methods (Scheirer 1996).

Spectral Decomposition, Musical Objects, and Spectralism

In a broader musical context, the spectral decomposition of music audio is the extraction of music objects (Schaeffer 1967). In this context, spectral decomposition is similar to the techniques used in spectral music in the sense that a computer analysis of audio-based features provides ways to extract compositional materials that could remain inaccessible to the composer (Dufourt 1979, Moscovich 1997). For the spectralists, this means extracting frequency information from audio samples using the fast fourier transform (FFT) (Brigham and Morrow 1967, Smith 2003) and the Short-Time Fourier Transform(STFT) (Arbo 2009). Composers Gérard Grisey, Tristan Murail, and Henry Dufort invented new harmonic structures using the extracted spectral profiles, often with novel tuning metrics (Grisey 2008, Arrell 2002). Their work reflects more than devising a new system of composition in the Schoenbergian sense; instead, it reflects a new way to perceive or experience music. This is shown by the careful attention to how frequency information forms the percept of timbre in many of the earlier works, and comments by Grisey in a 1996 interview echo this sentiment:

> Spectralism is not a system. It's not a system like serial music or even tonal music. It's an attitude. It considers sounds, not as dead objects that you can easily and arbitrarily permute in all directions, but as being like living objects with a birth, lifetime and death. (Bundler 1996).

The exclusive reliance on FFT and STFT, however, does not show any attempt to reveal pattern similarity in the time-frequency information. Explicitly, this point raises the issue of time-frequency separability. In certain circumstances, music audio information is separable. That is, if frequency information in the time-frequency domain does not overlap, then it is possible to achieve relatively good separation of sources using filtering. Another way audio can be separable is when temporal events in an audio mixture occur at discrete times and do not overlap temporally (e.g., having the ability to separate an eight-second audio clip into two parts that relate to different events in the audio). The majority of sound sequencing and editing software is designed for this purpose (e.g., Audacity, Logic Pro etc.) and has ways of mitigating the effect of splicing audio, a term that refers to early tape editing methods (Fasciano et al. 1995).

These examples are limited, however, because most real-world audio is not this simple–especially music–with many potential overlaps occurring as a result of even simple audio effects. For example, applying background noise to a recording with non-overlapping audio can greatly reduce the separability of the sound sources (Virag 1999). Therefore, the separability problem becomes more difficult when frequency information from different audio sources overlaps. This occurs in basic situations, such as instruments with pitch variance (e.g., anything without a fixed pitch or spectral profile) and when instrumental sources are mixed together with spectral overlapping (e.g., violin and flute in the same recording). In these situations, filtering will not satisfactorily separate the components (see 3 for a detailed discussion).

How is information separated in these situations? In short, correlations need to exist between different audio features in order to identify discrete sound sources. Some researchers (Casey and Westner 2000) looked for correlations between temporal (i.e., rhythmic) and spectral (i.e., pitch and timbre) patterns, described as independent spectral components. These independent spectral components extracted using spectral decomposition offer something that is not offered by spectral analysis using FFT or STFT: that is, access to a complex set of latent interactions that yield perceptible patterns and shapes across the entire spectrogram.

Observing Dynamics Between Correlated Time-Frequency Components

The following metaphor describes spectral decomposition for music audio: Imagine a spectrogram of a particular piece of recorded music is actually an infinitely large orchestra. The performers have instruments consisting of single sinusoids. While the performers cannot play more than one tone, they can adeptly shape the envelope of their sine waves. When these performers organize into small groups–much like sections of an orchestra–they can produce an expansive range of timbral possibilities.

If a person observed these different sections over time, he or she would witness consistent or emergent patterns of behavior and start to see that these behaviors are not only a result of their distinctiveness, but also a result of the dynamics of their behaviors over time. This type of analysis would reveal that there is not only structure in how the players are organized (i.e., the orchestration of the music), but also structure in how the music unfolds over time.

When the infinite orchestra metaphor is considered in the context of spectralism, then accessing interactions between sources in a spectral decomposition transforms spectralism into temporalism. This means the interaction of spectral structures (e.g., a single event and timbre) in a mixture of sources is as important to analysis as the spectral structures themselves. As Chapter 3 demonstrates, Grisey's s*Partiels Pour 16 Ou 18 Musiciens* explores the temporal interactions of features as much as using frequency information to build novel harmonies (Grisey 1975). Spectral decomposition, therefore, extends temporal interactions in spectralism to include the interaction of spectral structures.

Independent Spectral Components as Representations of Timbre

The concept of timbre is implicit in a musical interpretation of spectral decomposition. Risset and Wessel (1982) defined timbre as "a phenomenon consisting of both an average spectra (spectral profile) and how tones begin and end," or the amplitude-envelope information (pp. 114-115). This is the per-component information extracted in spectral decomposition. Specifically, in a non-negative matrix factorization (NMF) context, a spectrogram **X** forms as a consequence of the hidden matrices **H** and **W** (Smaragdis and Brown 2003):

$$\mathbf{X} = \mathbf{W}\mathbf{H}, \tag{1.1}$$

where **W** describes the spectral frequency profile information, and **H** is the amplitude-envelope information. Using the **k**th **W** and **H** column, it is possible to generate what is described here as a independent spectral component: x_k, defined by $x_k = w_k h_k^T$, respectively. These features, therefore, relate directly to timbre because they are similar to the time-varying amplitude and frequency functions extracted from a digital analysis of instrument tones (Risset and Wessel 1982).

Informing Temporal Similarity through Timbre

Independent spectral components offer a new way to explore the relationship between structures in spectrograms because they convey a better representation of timbre, but the particular design or application for spectral decomposition in music-related activities remains open to many unexplored possibilities.

The research discussed in this dissertation builds on recent approaches to spectral decomposition in music analysis by offering an approach to digital musicology, and a content-based retrieval design for rhythm pattern retrieval. While there are many different ways to analyze rhythmic patterns, the search by groove method (Topel and Casey 2011a, Qingyuan et al. 2011) is used to analyze rhythmic patterns in the present study. The concept of groove can be quite complex, but it is limited to what Hughes (2003) referred to as autotelic (p. 15) or self-generating rhythmic grooves, which can be repetitive or quasi-repetitive rhythm content. Butler (2006) suggested a simpler definition and defined groove as “the pattern laid down by the bass and drum kit” (p. 5).

The attractiveness of groove retrieval comes from the following intuition: groove often consists of stable and predictable timbres, often described as the rhythm section of a band or ensemble. In addition, these background rhythms can be shared across genres and provide an opportunity to examine music similarity on large datasets outside the genre paradigm. The contributions of this research include the following:

- Showcase of a collaborative groove retreival system.
- A publicly available dataset of 1,138 commercial rhythmic dance music tracks for rhythm retrieval experiments with a category-by-groove markup provided by two expert listeners.
- An evaluation of a groove retrieval task using the ISHKUR dataset and markup.

The present study has been conducted to examine whether latent structure can be accessed across an entire dataset using spectral decomposition and its relationship to the relatively invariant timbral qualities of grooves. To test this idea, the dataset ISHKUR, named after the popular online resource Ishkur's Guide to Electronic Music, comprised of 1,138 tracks, was given to two expert listeners who sorted the collection into what they thought were similar grooves. These markups were then used to evaluate a groove retrieval method using a novel timbre channel and group similarity algorithm collaboratively invented by Michael Casey and Qingyuan Kong (Qingyuan et al. 2011).

1.1.2 A Few Questions

In this study, spectral decomposition applications are explored beyond source separation in order to obtain a better understanding of structure in music audio and how to manipulate different components related to these structures in a variety of contexts. Therefore, the following questions guided this research:

- Although there are many references to auditory scene analysis (ASA) in the audio spectral decomposition literature (Casey and Westner 2000, Shao and Wang 2008, Shashanka 2008), how does ASA relate to spectral decomposition techniques?
- Much of the psychoacoustic literature identifies representations of listening, but how do such representations relate to algorithms designed to separate information in audio signals?
- To what extent does timbre influence structure within a musical signal? This question forms the basis for the theoretical framework discussed in Chapter 3.
- What can spectral decomposition tell us about a spectralist composition? Prior attempts at analyzing this corpus of music using traditional analysis techniques often avoid investigating the principle claim of these compositions, that they are derived from audio sources. If so, does spectral decomposition offer a potentially better way to analyze spectralist compositions?
- Given that current models of spectral decomposition are more than 10 years old (Casey and Westner 2000), what are possible future directions for these tools?

1.2 Dissertation Structure

The remaining chapters of this dissertation present the following arc: (a) a discussion about unconscious listening representation from an auditory scene analysis perspective (Bregman 1994); (b) a presentation of a latent structure framework and specific analysis examples from two pieces: *Partiels Pour 16 Ou 18*

Musiciens by Grard Grisey (1975) and John Cages (19401947) Fourth Interlude for Prepared Piano from the *Sonatas and Interludes*; (c) a discussion about a recent collaborative database study of rhythm retrieval using an extended version of probabilistic latent component analysis (PLCA) called Hierarchical-PLCA (H-PLCA); (d) a presentation of recent compositions using spectral decomposition methods; and (e) a discussion about the contributions and future possibilities of spectral decomposition.

Chapter 2 contains a discussion about the representation of unconscious or primitive listening from psychoacoustic and perception perspectives, and special attention is paid to the work of Albert Bregman and comments about listening made by Diana Deutsch and Carol Krumhansl. It also contains a discussion about how Gestalt principles influence our understanding of intermediate representation in cognition and its relationship to spectral decomposition.

Chapter 3 formalizes this concept using Pierre Schaeffer s (1947) ideas about objets sonores or sound objects and timbre, which has implicit connections to spectral music, a potential precursor to latent structure analysis using spectral decomposition first identified by Topel and Casey (2011b). Specifically, it is argued that spectral decomposition enables us to intercept the low-level auditory formulations of latent structure. Spectral decomposition is then applied to contrasting analyses, the first being an analysis of the two musical works described above. The second analysis, presented as a retrieval experiment in Chapter 4, explores the notion of groove as it relates to a generalization of timbral structure across a $1,187$ track dataset.

Chapter 5 contains a discussion about an approach to audio analysis that results in a range of possibilities for composition. In particular, the pre-audio and post-audio transformation of source-separated components are explored using compositions by five composers: (a) David Plans-Casal, (b) Michael Casey, (c) Simon Atkinson, (d) Paul Osentinsky, and (e) Spencer Topel.

Chapter 6 contains a discussion about the general contributions and implications of the present study. Three additional sections are included as appendices: (a) a glossary of abbreviated terms, (b) preliminary experiments using spectral decomposition to retrieve rhythms, and (c) complete examples of the Matlab code paraphrased in the main body of this dissertation.

1.3 Collaborative Nature of Latent Structure Analysis

The analysis in chapter 4 is part of a larger study conducted by the Bregman Music Audio Research Studio (BMARS). While this analysis was the initial focus of this dissertation, for a variety of reasons, it eventually became a section in a broader creative and research effort. The other analysis was performed by the author of this dissertation in an environment driven by collaboration with faculty and graduate students at Dartmouth, and I have been careful to acknowledge other people's experiments, ideas, and analyses.

The compositions listed in chapter 5 illustrate a seminal trend toward works created explicitly using source-separation algorithms, and they do not represent the entire output of possible works. In fact, since this this dissertation was written, new works have been created that could certainly be added to the small set of works presented in this study.

1.4 Code, Music, and Example Guidelines

All code examples in this dissertation appear as Matlab scripts and functions. This enables interested researchers in different music-related areas, such as music information research, signal processing, and computational perception, to quickly audition examples and functions. In addition, code examples are kept to a minimum in the main body of the text, with longer and more complete sections of code provided in Appendix B.

The syntax of the Matlab code follows similar conventions found in other programming languages. In Matlab, comments are delineated by the % symbol and loops follow the "C" indentation guidelines. Functions embedded in parent functions (as abstractions) appear at the end of the parent function and are always terminated by the "C"-style "Return" command.

Examples in the main chapters are designed to be instructive, and there are comments that provide details about step-by-step processes. Executable examples with supporting functions can be found at the Bregman Music and Audio Research Studio (BMARS) (for information about where to access these examples, see (Topel 2012a).

Music examples are presented in two forms: (a) Western music notation and (b) music features consisting of spectrograms, waveforms, histograms, and components extracted from PLCA. For the few cases in which extended Western notation is employed, a key defining the notations is provided. Audio examples that correspond to all the music examples in this study can be found online (see Topel (2012b)) and are labeled by the music example's figure number (e.g., Figure 5.1).

CHAPTER 2

REPRESENTATION OF UNCONSCIOUS LISTENING IN MUSIC

> *Music presents us with a complex, rapidly changing acoustic spectrum, often resulting from the superposition of sounds from many different sources. The primary task that our auditory system has to perform is to interpret this spectrum in terms of the behavior of external objects.* Deutsch (1982).

Deutsch (1982) stated if all [acoustic] first-order elements were indiscriminately linked together, auditory shape-recognition operations could not be performed (p.100). Krumhansl (1992) similarly warned against thinking of perceived or imagined music as auditory tapes in the head which record sound-pressure variations continuously over time, but rather, music is organized at early levels of processing into events, properties of events, and temporal relations between events (pp. 200-201).

The process by which listeners arrive at the shape-recognition discussed by Krumhansl and Deutsch suggests that there are correlated features in sounded music. The investigation of these correlated behaviors forms the basis of Bregmans (1994) research that investigated ASA. He suggested listeners perform complex scene separation processing, which results in stream segregation, in which low-level auditory observations fuse to form a sense of continuation and a single percept. Bregman related the process governing streaming to a principle in Gestalt psychology known as common fate, which is the sense of inevitability created by correlated audio features.

The present study is interested in primitive segregation, which is thought to encompass the low-level auditory decomposition of a scene and arises from a bottom-up strategy of parsing information based on correlations and auditory cues or events. Although it has been shown (Zendel and Alain 2009) that low-level auditory decomposition improves through practice, it is a skill that all humans and some animals possess (Bee and Micheyl 2008).

Primitive segregation contrasts with what Bregman (1994) called schema-based segregation. Intuitively, schema-based segregation describes a top-down approach to listening, in which there is an effort to discern different patterns, or what is informally called active listening (Okamoto et al. 2010). The suggestion that certain types of percepts arise out of one category or the other or both (LaBerge 2010) does not add anything to the present study and is not used in analysis. Primitive segregation is important for the present study because it describes a latent cognitive behavior, something listeners do unconsciously. This unconscious processing attribute of primitive segregation makes streaming, in Bregmans (1994) words, “not a simple thing to measure” (p. 54).

2.1 Representations of Listening

At the most basic level of representation, all listeners have to deal with the same problem: Sounds that occur simultaneously in a surrounding environment reach our ears as a single pressure wave (Bregman, 1994, p 107). These concurrent and continuous pressure waves can be though of as audio or auditory mixtures. Bregman identified two processes people use to listen: (a) Listeners rely on schemata or prior auditory knowledge to separate mixtures, and (b) listeners employ a primitive process to group incoming electro-physical impulses. He defined *primitive process* as “a multi-dimensional decomposition of time events, and frequency events, resulting in the continuous observation of multiple streams" for the following types of information:

- intensity
- fluctuation patterns
- direction of frequency transitions
- estimation of sound source locality. (p. 613)

2.1.1 Anatomy of Auditory Scenes

Using two strategies simultaneously, humans possess the ability not only to enjoy music, but also to have conversations with another person in a noisy space, operate moving vehicles, and perform all types of tasks that involve separating important sound cues from everything else. Bregman (1994) suggested that auditory streams exist in mixtures called auditory scenes.

Figure 2.1 illustrates ASA using a spectrogram to represent the auditory decomposition problem. The top spectrogram shows easily recognizable patterns in the spectrogram that correspond to the spoke phrase "one, two, three." The bottom spectrogram, however, shows a mixture of different sources, which obscure the original pattern. It is the job of ASA to decompose the sources into their constituent parts, and this happens because a listener can easily distinguish sources from one another.

Bregman et al. (1971) argued that mixtures of sound sources in environments where different sounds overlap in time and frequency require a listener to perform auditory decomposition. This is a process in which "correlated amplitude

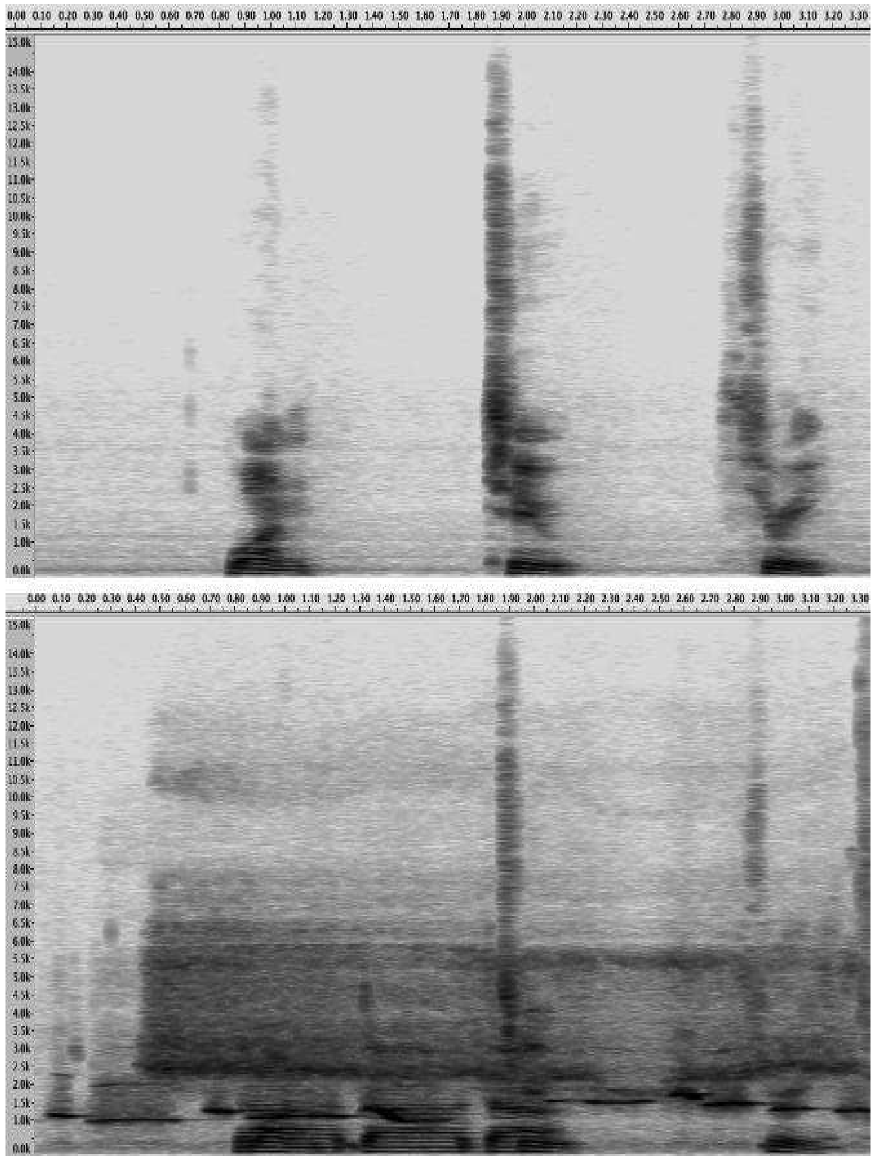


Figure 2.1: A reproduction of Bregman's classic example of auditory scenes. The top spectrogram has the spoken phrase one, two, three, and the bottom spectrogram is a mixture of (a) the spoken phrase one, two, three, (b) singing da-da-da, (c) whistling, and (d) a computer fan (Bregman 2005).

changes in different parts of the spectrum contributes to the assignment of the spectral components to a perceived source" (Bregman, 1994, p. 590).

2.2 Streams and Streaming of Auditory Information

With respect to auditory scenes, the auditory system quickly groups various components within an incoming signal (Bregman, 1994). Components that are similarly grouped can be described as auditory streams. There are different opinions about the use of the term streams in relation to auditory perception. For instance, Deutsch (1982) referred to streams as any situation in which auditory stimuli separate into discretely observable parts, and listeners separate streams on the basis of sound type (p. 124). This use is different from the concept of streaming that suggests different perceived qualities in the sound come together spontaneously to create the auditory object (Bregman, 1994, p.10). The commonality between streams and streaming is in the grouping of similar qualities:

> I view a stream as a computational stage on the way to the full description of...auditory event(s). The stream serves the purpose of clustering related qualities. By doing so, it acts as a center for our description of...acoustic event(s). (Bregman, 1994, p. 10)

The important element in the quote above is "clustering related qualities." According to Bregman (1994), these are qualities that lead to a formulation of distinctive elements in an auditory scene, which is similar to the Gestalt principles of grouping and organization. Figure 2.2 illustrates a central concept in ASA, namely, that tones close in frequency (pitch) are more robust to streaming regardless of presentation order. When the distance between the frequency of tones increases, the tones form independent streams (Bey and McAdams 2003).

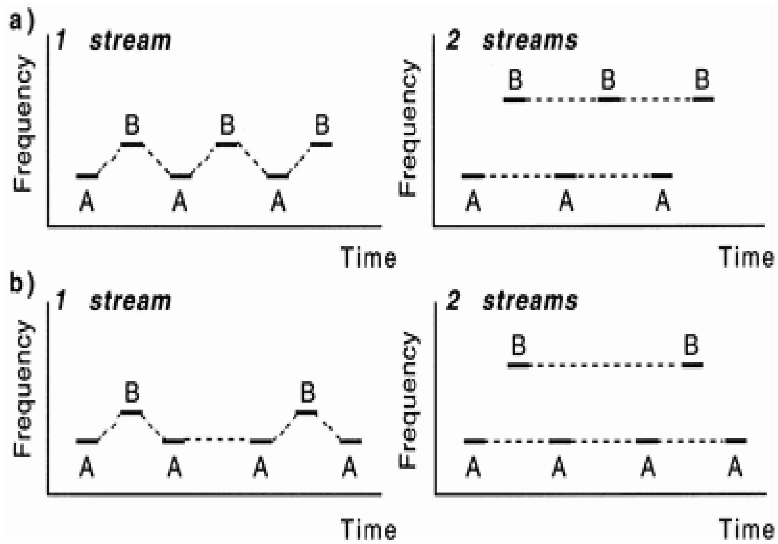


Figure 2.2: Tone-proximity streaming using auditory scene analysis (Bey and McAdams 2003)

The Gestalt principle of organization more broadly explains the perceptual formulations of streaming in ASA.

Gestalt Principles in Listening

Bregman's (1971) ideas about auditory scenes are largely based on the principles of perceptual grouping proposed by Gestalt psychologists, who suggested that, in effect, the interpretation of our world is a result of clustering similar things based on finite elements (Wertheimer 1938). Figure 2.3 illustrates three of the four primary Gestalt principles (Deutsch, 1982). The fourth gestalt principle, common fate, must be inferred and proposes that objects that share a similar trajectory are grouped together.

Deutsch (1982) stated that grouping by timbre is a result of the principle of Similarity (p. 124) and based on two factors: (a) the organization of a sequence

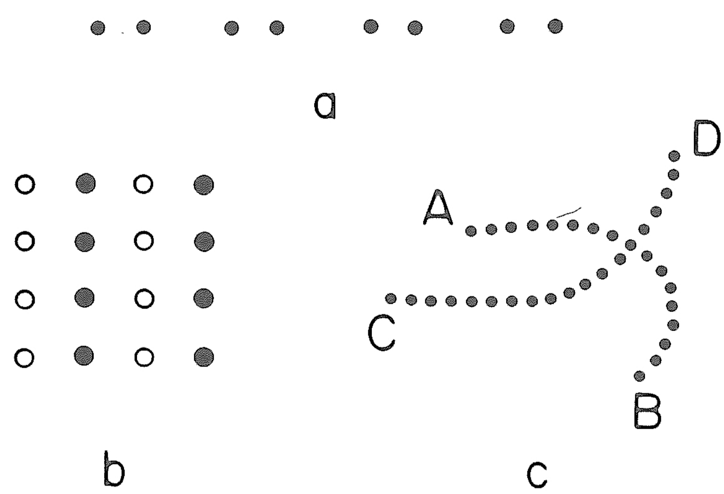


Figure 2.3: Graphical representations of the Gestalt principles of (a) proximity, (b) similarity, and (c) good form/continuation. Reprinted in "Grouping Mechanisms in Music," by Diana Deutsch, 1982, The Psychology of Music, 2, pp. 299348.

into separate streams on the basis of sound type and (b) familiarity with such sources (p. 125). The sequential groups discussed by Deutsch can be linked to the idea of timbre similarity as it relates to the unfolding of temporal events relative to a length (e.g., temporal segments) and auditory stream segregation (e.g., independent spectral features attributable to timbre) belonging to an auditory scene.

In contrast, Bregman (1994) identified common fate as the most predominant Gestalt axiom in auditory scene formulation. He posited that it is only through the linkages of behavior through time that listeners can determine if frequency information conveys any sense of structure. Bregman suggested that similarity, particularly with respect to familiarity of sources, belongs to auditory selective attention processes and not exclusively to primary organization.

This is an important distinction because it allows for competing interpretations of auditory patterns. Gestalt psychologists argued that innate knowledge can be overridden when known patterns are confused with unknown patterns (Bregman, 1994). In a more recent study (Serafini 2010), researchers found that musicians trained in Javanese Gamelan classical music had different listening strategies with respect to low-level features than musicians trained in Western classical music, which suggests a bifurcation in auditory processing.

2.3 Computational Modeling of Auditory Grouping

An important component of ASA is the distinction between sequential integration, or how sounds unfold over time and their relationships, and simultaneous integration, or perceptual fusion, which describes how sounds decompose when they happen at the same time (Bregman, 1994). These two main strategies (i.e., simultaneous and sequential integration) are the basis for devising vertical and horizontal musical processing from low-level features, respectively (Rosenthal and Okuno 1998, Brown and Cooke 1994, Lerdahl et al. 1996, Deliege 1987).

While these efforts at modeling human auditory perception are novel, in a 1998 critique, Slaney discussed a concept called pure audition, which suggests that inherently bottom-up approaches to computational auditory scene analysis (CASA) might "ignore the avalanche of information from higher cognitive levels" (Slaney 1998; p.28). More recently, Wang (2005) suggested that some type of perfect ASA is unrealistic and a target signal MIR retrieval paradigm is more realistic and testable. An alternative to elaborate models such as CASA and other models is a model that examines the latent structure in music audio with minimal assumptions and computational stages. For the purpose of this dissertation, the identified method for achieving low-level CASA stems from an algorithm called PLCA that uses a probabilistic framework (Shashanka et al. 2007). Specifically, this algorithm uses a generalization of Hoffmans probabilistic latent semantic analysis (PLSA) (Hoffman 1999), which extended the applicability of the latent semantic indexing (LSI) singular value decomposition (SVD) mode (Deerwester et al. 1990).

2.3.1 Spectral Decomposition in the Context of ASA

Hoffman (1999) achieved the PLSA extension to LSI by proposing a probabilistic bivariate statistically correct generative model based on the likelihood principle. Applying this same approach to spectral-audio information yields similar information using the likelihood function and accounts for what might be contained in spectral information without over-determination. The link between latent structure analysis using probabilistic spectral decomposition and ASA can be distilled to three important concepts:

- **Music audio has an inherent structure:** This can be applied to both high- and low-level structures, such as segmentation of musical sections to timbre, which is an outgrowth of the central premise in text-based latent semantic indexing.[1]
- **Extracted latent components satisfy simultaneous and sequential integration conditions for primitive segregation:** Bregman (1994) suggested

[1] The authors argued that LSI is possible because documents and collections of documents contain inherent structure, which can be discovered with minimal assumptions (Deerwester et al. 1990).

that while these two dynamics contribute to the separation of auditory patterns they are not separate processes. Correlated amplitude envelope and frequency profile information solve this problem by weighting the relative contribution of any one component to the mixture (Raj and Smaragdis 2005)

- **A priori information folds top-down inference into the core computation**: Instead of separating a grouping from scene decomposition (i.e., through clustering features), spectral decomposition used in this dissertation accommodates known data by training the basis functions used to estimate source content (Qingyuan et al. 2011) or structure (Weiss and Bello 2010).

2.3.2 Intermediate Representations in Unconscious Listening

Given the interrelated properties of ASA and probabilistic spectral decomposition methods, how may spectral decomposition situate itself in ASA? Figure 2.4 illustrates the spatial relationship between ASA and spectral decomposition. The left side of the figure illustrates the cognitive space of music (e.g., gist, thought, and abstraction). The right side of the figure illustrates the physical or substantive experience of music, such as performance and listening (Madsen et al. 1993). The inward facing arrows indicate Bregmans (1992) concept of high-level schema-based segregation and low-level primitive segregation. The dotted arrow next to the spectral decomposition space describes the rank estimation, or the number of components extracted using spectral decomposition. The idea is that as the number of components in a given extraction increases from some unknown IDEAL, the extraction parameter approaches the upper

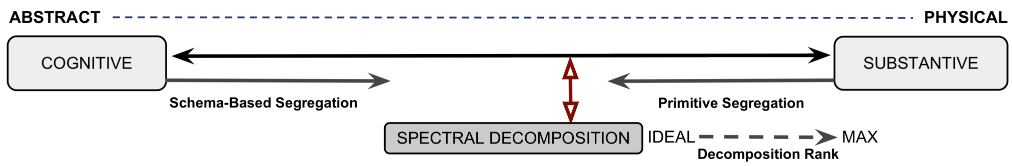


Figure 2.4: A diagram of ASA in relation to spectral decomposition, with the left side of the figure showing cognitive space, where music is in abstract forms, such as ideas or gists (Agres and Krumhansl 2008), and the right side is the physical world, where music is sound. It is the task of ASA to translate auditory information from the real world into abstract cognitive representations.

limit MAX, defined by the number spectral bins in the spectral analysis. As shown in this figure, the extraction of independent spectral components starts to resemble sinusoidal components and eventually becomes indistinguishable from the original spectrogram.

To illustrate the connection between ASA and spectral decomposition into independent components, consider Figure 2.5, which shows the phrase "one, two, three" approximated from the auditory scene example in Figure 2.1. Using a twenty-component PLCA extraction, the extraction satisfactorily extracts independent patterns relating to the sources. Thus, spectral decomposition rests somewhere closer to substantive experience in this model because the primary function of extracting independent spectral components coarsely mimics low-level primitive segregation (Elhilali and Shamma 2006). This is because spectral decomposition in its most basic form is agnostic to sources within a mixture and makes no attempt to group spectral patterns beyond the latent correlations in the data. This presents a potentially useful function for supervised activities, such as composing or remixing music (Topel and Casey 2011b) and a conversely

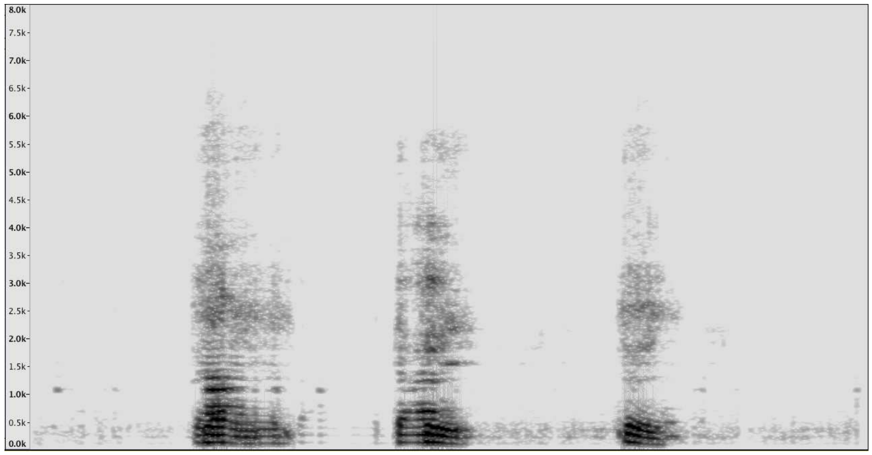


Figure 2.5: A twenty-component PLCA extraction recovered the spoken phrase one, two, three, from the Bregman Auditory mixture shown in Figure 2.1

problematic situation for unsupervised activities such as retrieval over many tracks in a dataset (Cardoso 1989).

The ability to access important structural aspects of sound by intercepting the intermediate stages of ASA before the cognitive formulation of the auditory scene is the underlying concept of latent structure analysis using probabilistic spectral decomposition, and it is similar to CASA (Wang and Brown 2006). This is important when one considers the possible applications of spectral decomposition for the analysis of music and creativity and, more important, the context of spectrally decomposed components with respect to listening.

2.4 Summary

This chapter presented representations of unconscious listening interpreted using ASA. This discussion provides the context for spectral decomposition as it relates to detail, such as timbre, similarity, and information channels, and higher-level concerns, such as composition, analysis, and retrieval.

This chapter contends that spectral decomposition in a probabilistic framework intercepts a lower stage of ASA without making any assumptions about grouping or similarity beyond the latent correlations in the information. This is a consequence of the following connections between probabilistic spectral decomposition and ASA:

- Music audio has inherent structure
- Correlated time-frequency components satisfy the simultaneous and sequential integration conditions for primitive segregation.
- A priori information folds top-down inference into the core computation

The investigation of latent structure throughout the remainder of this dissertation focuses on low-level primary segregation and how this type of information expands our understanding of music. It is arguable that some of the subsequent analysis points to high-level cognitive inference, particularly with respect to timbre (Wedin and Goude 1972), but while it might be possible to model active listening, it is beyond the scope of this research.

CHAPTER 3

TOWARDS A THEORETICAL FRAMEWORK FOR LATENT STRUCTURE IN MUSIC

This chapter examines how spectral decomposition can be used to formulate new interpretations of music by accessing sub-perceptual information hidden to listeners.[1] By design, spectral decomposition algorithms discussed in this dissertation maximize the independence of spectral features. This process is similar to primitive segregation, and it is argued that spectral decomposition using independent features is a perceptually informed process (Topel and Casey 2011b) similar to Bregman's (1994) ASA. It is also argued that if latent patterns could be consistently recovered, then it would be possible to develop new interpretations of musical structure.

3.1 Spectral Analysis in Music Composition

The application of spectral decomposition methods for music composition, extracted most commonly using FFT or STFT, has an extensive history in computer-assisted music composition genres: notably (a) Wisharts expansion of Pierre Schaeffers Music Objects (Wishart and Emmerson 1996) and (b) Spectromorphology proposed by Smalley (1997) . The second half of this chapter considers spectral decomposition as a means to answering musicological questions. Two compositions are analyzed: (a) Interlude four No. 4 from *Sonatas and Interludes for Prepared Piano* (1946—1948) by John Cage and (b) Gérard Grisey's

[1]Unconscious listening describes only the formulation of sound cues into patterns and not the agent-based models found in traditional neuroscience literature (e.g., McGuiness and Overy (2011)).

(1975) *Partiels Pour 16 Ou 18 Musiciens*. They were chosen for two reasons: (a) both contain explicit timbral transformations of their instruments or ensembles and (b) the means by which both composers derive their material are so different. Cage, for instance, arrived at the preparations for the pian through intuition and listening. In contrast, Grisey used computer analysis to uncover specific frequencies and amplitudes related to the timbre of the source material. This kind of computer analysis was a new concept at the time and forms the basis of spectralism.

3.1.1 Spectrale Musique

Henry Dufourt introduced the idea of spectrale musique in 1979; however, by this time, compositions using material extracted using the Fourier transform had already been written by the group, notably Gérard Griseys (1975) Partiels. In most of these early pieces, there is a transparent, straightforward process used by the composer (see Figure 3.1).

Following this research and these compositional activities, a well-spring of software development at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) occurred, most notably the visual programming environment Max, after Max Matthews. In addition, many of the well-known IRCAM packages were developed: (a) Music V, brought to IRCAM by Jean-Claude Risset; (b) *CHANT*, developed by Xavier Rodet for the purpose of formant analysis; and (c) the transcription tools that now belong to Open Music. Grard Grisey, Tristan Murail, Hugues Dufourt, and British composer Jonathan Harvey were part of the first wave of spectral composers. Their music influ-

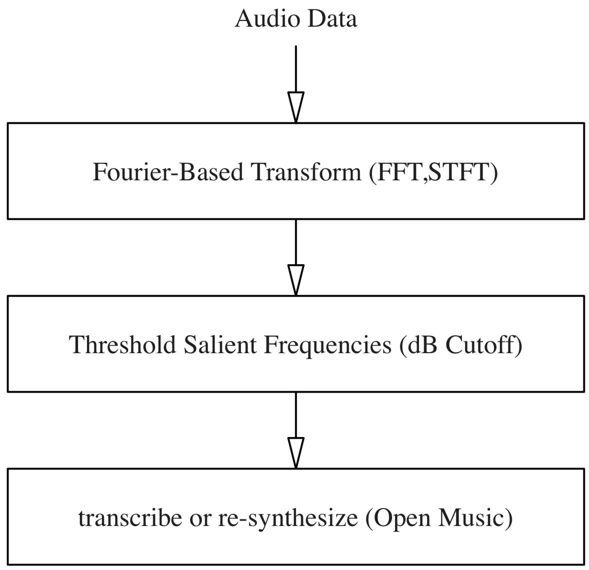


Figure 3.1: Typical analysis procedure in Spectrale Musique: A relatively simple method, which would yield harmonicically novel material. The first step would be to perform a Fourier transform, that either extracted only frequency information (FFT), or preserved temporal resolution of the frequencies (STFT). The former technique was employed in the earliest compositions on relatively simple audio samples, (e.g. ringing bell and a single piano note).

enced a younger cadre of composers, including Magnus Lindberg, Marc-Andre Dalbavie, and Joshua Fineberg.

This analysis has deep connections to the European *Zeitgeist* of the 1960s and 1970s. At that time, total serialism was considered the only credible method for Western-music based composers. Total serialism emerged as an approach to music composition inspired by the post-war generation of composers. Namely, Pierre Boulez, Karlheinz Stockhausen, under the auspices of the Darmstadt summer school (Grant 2005).

The spectralists realized the aesthetics associated with the mainstream serialism of the time disregarded the final sounded musical experience. Instead, these techniques favored abstraction in notation and formalism, and composers were blinded, to a certain extent, by the idea that all 12-tone (half-step) relationships were equal and non-Western tuning systems were outside the scope of mainstream Western art music. Instead, the spectralists considered latent frequency information from acoustic sources to be the starting point for their work. They further rejected the idea that the 12 chromatic tones were equal and non-traditional tuning clashes with Western tuning. Their early repertoire emphasized these points, with pieces such as *Gondwana* (Murial 1980), which evokes a sense of non-Western tuning by exploring the time-frequency relationships between a synthetic bell source and a trombone sample. Later, *Désintégrations* (Murail 1989) used the careful blending of timbres from electronic sounds and acoustic sources.

3.2 Revealing Latent Structure in Music Audio

As discussed in chapter 1, latent structure refers to distinctive or salient parts of recorded audio that otherwise remain hidden to the listener. For the spectralists, this means identifying structural partials that distinguish one instrument from another instrument playing the same perceived note (e.g., an A or Bb). With concatenative synthesis (see Chapter 6), the latent structure emerges from the corpus through the similarity of audio features.

Independent component extraction techniques offer yet another way to access structure in audio because resynthesized components retain correlated behaviors between frequency and amplitude information in each component. When PLCA is used on magnitude-only STFT representations, the extracted components have characteristics similar to the output of phase vocoder methods. There is, however, an important distinction: In addition to spectrum and envelope decompositions, components are further segmented by the independence of information content or patterns (Bailey et al. 1994).

3.2.1 Probabilistic Latent Component Analysis (PLCA)

The decomposition of an audio signal requires some model that can find independence in the features. The primary algorithm used to process audio component extraction in this dissertation is the PLCA algorithm (Smaragdis et al. 2008). In general, PLCA expands non-negative matrix factorization (NMF) by introducing a probability framework. In NMF, a non-negative matrix **V** is decomposed into the product of two matrices **W** and **H**:

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \tag{3.1}$$

where **V** is a time-frequency decomposition of an audio signal, such as a magnitude spectrogram, each column of **W** is a frequency signature corresponding to frequency characteristics, and each row of **H** is the

temporal activations of the extracted components (Weiss and Bello 2010). PLCA recasts NMF in a probabilistic framework, and the resulting equation can be written in NMF terms:

$$\mathbf{V} \approx \mathbf{WH} = \sum_{k=0}^{K-1} w_k z_k h_k^T \tag{3.2}$$

Using the expectation maximization (EM) algorithm (Moon 1996), the posterior distribution over latent variables is computed for each cell in $\mathbf{V}$ in the E-Step, followed by the maximization of the parameters using the posterior distribution calculated in the E-Step. The updated parameters corresponding to w_k, z_k, and h_k are then iteratively computed through EM until a convergent solution is obtained.Figure 3.2 illustrates the decomposition of magnitude STFT time-frequency distributions into independent components using this two-dimensional marginal decomposition algorithm.

For procedures aimed at generating composition material, the **Z** Prior can be modified in re-synthesis either by omission or normalization (e.g. by setting z to a constant variable) because it functions as the relative contribution (or loudness) of each component to the original mixture.

The usage of PLCA can also be applied to extraction of notes from instrumental lines. Figure 3.3 is the code used to extract and resynthesize the audio waves shown in Figure 3.4. In this context, each component is a single pitch-class extracted from an audio recording of the first fourteen seconds of the *Chaconne* from *Partita No. 2 in D minor for solo violin*, by J.S. Bach (BWV 1001).

3.2.2 Parameter Selection in Spectral Decomposition

In most computer algorithms, there is an attribute called an open parameter that can significantly influence or change the resulting outputs (Aho and Hopcroft

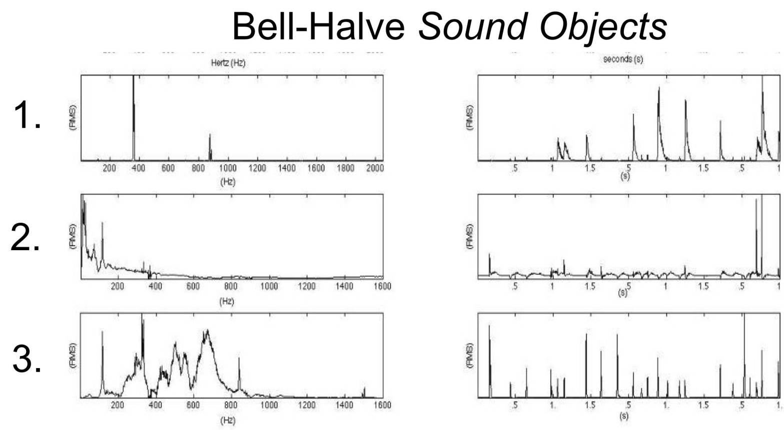


Figure 3.2: Components can be "sound objects" in the Schaefferian sense, when they contain recognizable yet independent traits of an instrument or sound. 1) the metal attack caused by the striking of cement 2) the ringing of the metal after excitation. 3) Oscillation or "wobbling" of the bell as it comes to rest on the cement.

1974). Open or free parameters can generally be described as user-controlled variables. They control aspects of the algorithmic function such as window size, analysis resolution, and component estimation. For instance, if the bin size in an FFT is relatively large, then the resulting representation contains less frequency resolution. If the FFT window size (over the audio data) is large, temporal resolution is diminished while maximizing frequency resolution. For scientific applications, determining the correct open parameters is often a matter of performance. In other words, the best-performing parameters are the correct ones.

```
% Load Bach Chaconne excerpt into Matlab
s = wavread( 'chaconne.wav')';

%sum channels to produce Mono signal
s = s(1,:) + s(2,:);

% Go to time-freq domain via STFT
f = stft( s, 4096, 1024, 0, 'hann');

% Perform PLCA analysis on eight channels
[w,h,z] = plca2d( abs( f), 8, 200, 0, 0, 0, [], [], [], 1);

% Resynthesize components to audio iteratively
fn = abs( f) ./ (w*diag(z)*h);
fp = f ./ abs( f);
fp = fp ./ repmat( linspace( .5, 10, size( f, 1))', 1, size( f, 2));

for i = 1:size( w, 2)
    tf = (w(:,i)*z(i)*h(i,:)) .* fn;
    y(i,:) = stft( tf.*fp, 4096, 1024, 0, 'hann');
    wavwrite(y(i,:), 44100, [filestem '_' num2str(i, '%02d') ...
        '.wav']);
end
```

Figure 3.3: A Matlab script performing an eight component PLCA decomposition on the opening of the Bach Chaconne, using the PLCA2d function proposed in (Shashanka et al. 2007).

For composition, selecting the right parameters is not straightforward. It is also more flexible in terms of how they can be applied and evaluated. In the case of independent component analysis, for instance, component estimation becomes an expressive parameter. Over-estimation or under-estimation of components for a given mixture becomes a compositional question specific to a work-in-progress rather than a means for successfully separating sources. In addition to open parameters, the selection of the appropriate re-synthesis method greatly affects the result, a topic explored in chapter 5.

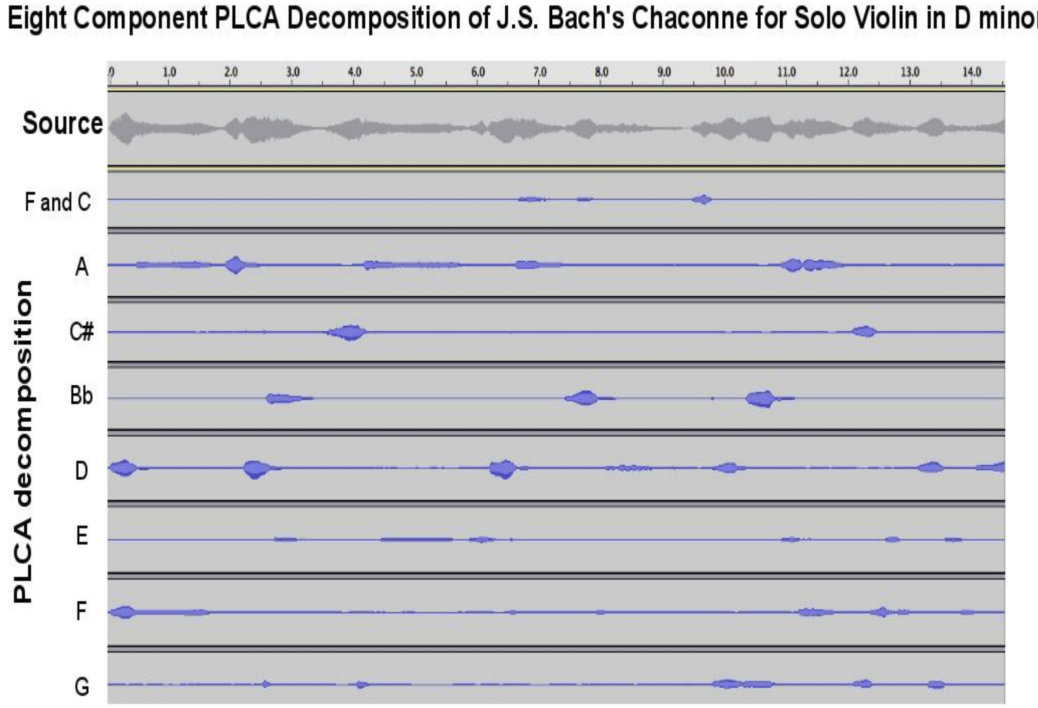


Figure 3.4: Components extracted from a performance of the first phrase (approx. mm. 1-4) of the Bach Chaconne for solo violin. In this case, the components correspond to specific notes or frequencies that reoccur, shown here as resynthesized waveforms

3.3 A Theoretical Basis for Latent Structure Analysis

The following section discusses possible outcomes of spectral decomposition and their applications. When these techniques are used on music spectral features (e.g., STFT matrices), the basic decompositions yield components similar to those hypothesized by Schaeffer (1967) in his discussion about sound objects, music objects, and magnets. However, this comparison alone accounts for only another way to re-combine timbral elements, a reverse-engineering approach of modeling musical instruments presented as a synthesis modeling problem by McIntyre et al. (1983).

The characteristic that best identifies spectral decomposition for music processing is the preservation of timing and frequency information for entire music-audio scenes. This means relationships between different musical events can be decomposed and recombined while preserving, in essence, basic features that define the specific relationships between sounds and events that make them recognizable, that is, the latent structural attributes of music audio.

3.3.1 Component Extraction as Timbral Decomposition

Osetinsky (2010) suggested that spectral decomposition and granular synthesis are related to one another. Although the analysis window in spectral decomposition algorithms can behave like a grain window by setting the parameters to very small values (durations), the processing that occurs in a grain window is usually quite simple, often consisting of an envelope-shaping kernel that affects the attack and decay of the grain. This has no relationship to the procedure in spectral decomposition because the feature or information itself is responsible for the shapes of the spectral profiles and amplitude envelopes.

Granular synthesis and PLCA, however, are similar in relation to the semantic attribution of the process output, outlined in the brilliantly well- conceived study on emotion attribution by Scherer and Oshinsky (1977). Semantic attribution suggests that some resynthesized components have clear or partially recognizable emotional and associative attributes, which under certain conditions can result in listener source identification (Lakatos et al. 1997). Other components are less recognizable or intuitively connected to the source but may still play a structural role in the analyzed audio sample, such as separated formants or other noise features, and these, in turn, lend themselves to a more-complicated interpretation of timbre (Erickson 1975).

Schaeffers (1967) notions about sound objects and musical objects are a closer analog to PLCA components than granular synthesis. In this treatise, Schaeffer described sound objects as identifiable atomic units of a musical object, and the musical object can possess one or multiple sound objects, or what is referred to as a source in MIR literature.

Schaeffer (1967) uses the example of a snare hit. He pointed out that the sound of the snare can be decomposed into a two different sound objects: (a) the noise of the hit and (b) the resonance of the drum. The composer went further and said dividing the spectrum in half using high-pass and low-pass filters results in musical objects that are still identifiable as the original musical object, or what he refers to as magnets . In a sense, Schaeffer (1967) is describing a timbral decomposition in which musical objects are the audio material, sound objects are timbral components possessing independent spectral characteristics, and magnets are components that retain enough characteristics of the original musical object to be mistaken for the original. This is the result of one of two primary factors: (a) features were unsuccessfully decomposed and resulted in hybrid components or (b) the audio material has features that are not independent and completely unified in their behaviors.

In the latter situation, the only way to separate the sources in the audio would be to use a deconvolution method with audio taken from an exact or random source (Douglas et al. 1997) or, conversely, synthesize the components, which forms the basis of Smalleys (1997) spectromorphology. In the case of the former, modifying the parameters of the component extraction algorithm may improve the analysis and yield a better decomposition.

Returning to the idea of sound objects, Figure 3.2 shows how PLCA components can be sound objects. This example is a PLCA decomposition of a brass bell-halve consisting of three identifiable sound objects: (a) the metal attack caused by striking cement, (b) the ringing of the metal after excitation, and (c) the oscillation or wobbling of the bell as it comes to rest on cement. When combined together, these components accurately reproduce the sounds recognizable as the bell-halves.

3.3.2 Sub-Mixtures in Music Audio

Compositions with multiple instruments or singers have different issues than the simple compositions described above. While single-instrument decompositions can be thought of as timbral decompositions, when there are more instruments or sounds that contribute to similar timbres, there are more shared components than with the former example. In such cases, it is not possible to perform perfect source separation (i.e., to extract every instrument or vocal part) using these methods. Yet, with techniques such as PLCA, it is possible to acquire consistent timbre information.

An example of timbral separation can be observed in Figure 3.5, where an audio excerpt containing a sustained dominant seventh chord and a rhythmic pattern played by a standard drum set are separated into two distinctive, consistent patterns of information that are shown in three representations: (a) log spectrograms, (b) PLCA H, W components, and (c) the resynthesized waveforms. Observe that each pattern contains traits identifiable as either sustained harmonic or rhythmic information. The sustained harmonic chord is shown on the top, and it has clear, even-spaced partials in the spectral features and extended noise-like energy in the amplitude features. Drum set rhythmic content is shown at the bottom. The spectral features show banded noise-like characteristics, and the amplitude features exhibit spikes of non-sustained energy across the window. In this relatively simple example, the separation into two timbres is successful because the patterns between the two timbres are sufficiently distinct or independent.

In more complex audio mixtures, where patterns are shared across multiple sources or there is variance between patterns, PLCA2d will be less effective in the extraction of the timbres. Figure 3.3.2 shows an example of a PLCA timbres extraction using best parameters that is similar to Figure 3.5. This time, however, the chords from Figure 3.5 are performed with short rhythmic articulations, and greater distortion is added to mimic the timbre of the drums. The resulting extraction creates two hybrid timbres that reflect an unsuccessful separation of the two timbral sources. In a sense, the harmonic-sustained chords have become harmonic-percussive, and they have more qualities in common with the percussive timbres in the mixture.

The mixture of timbres in this PLCA2d decomposition shows that close timbres are more likely to be inseparable. This raises the question of independence of timbres in a mixture and their separability as it relates to a concept Bregman (1994) called fusing. Timbres can fuse and become an ensemble of timbres when "subgroups of instruments are yoked together by the fact that their parts move in real parallel motion and their notes go on and off together because of an exactly duplicated rhythm" (Bregman 1994; p.520).

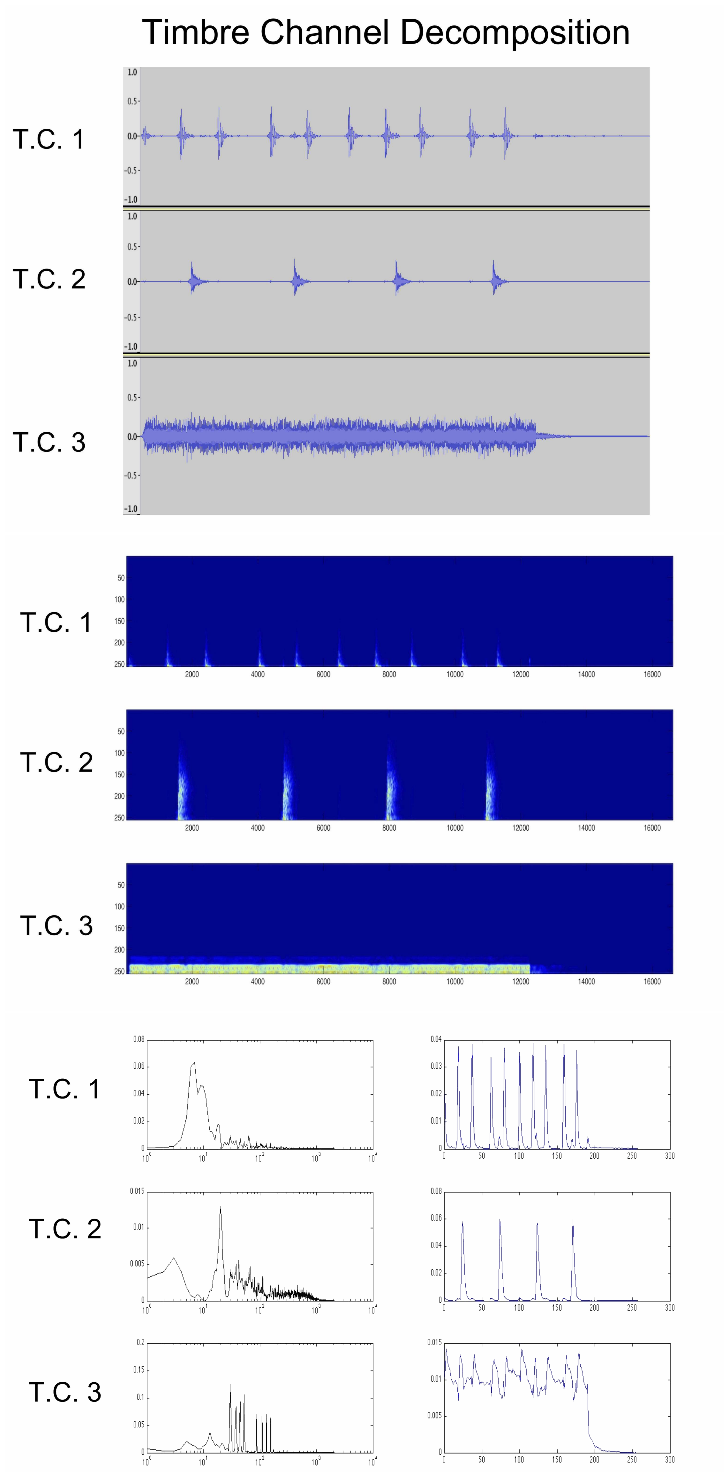


Figure 3.5: Ideal Timbre Extraction: where the snare (top), kick drum(middle), and sustained harmonic have been extracted into three discrete patterns and displayed as the following representations: resynthesized waveforms,log-spectrograms, and PLCA H, W components.

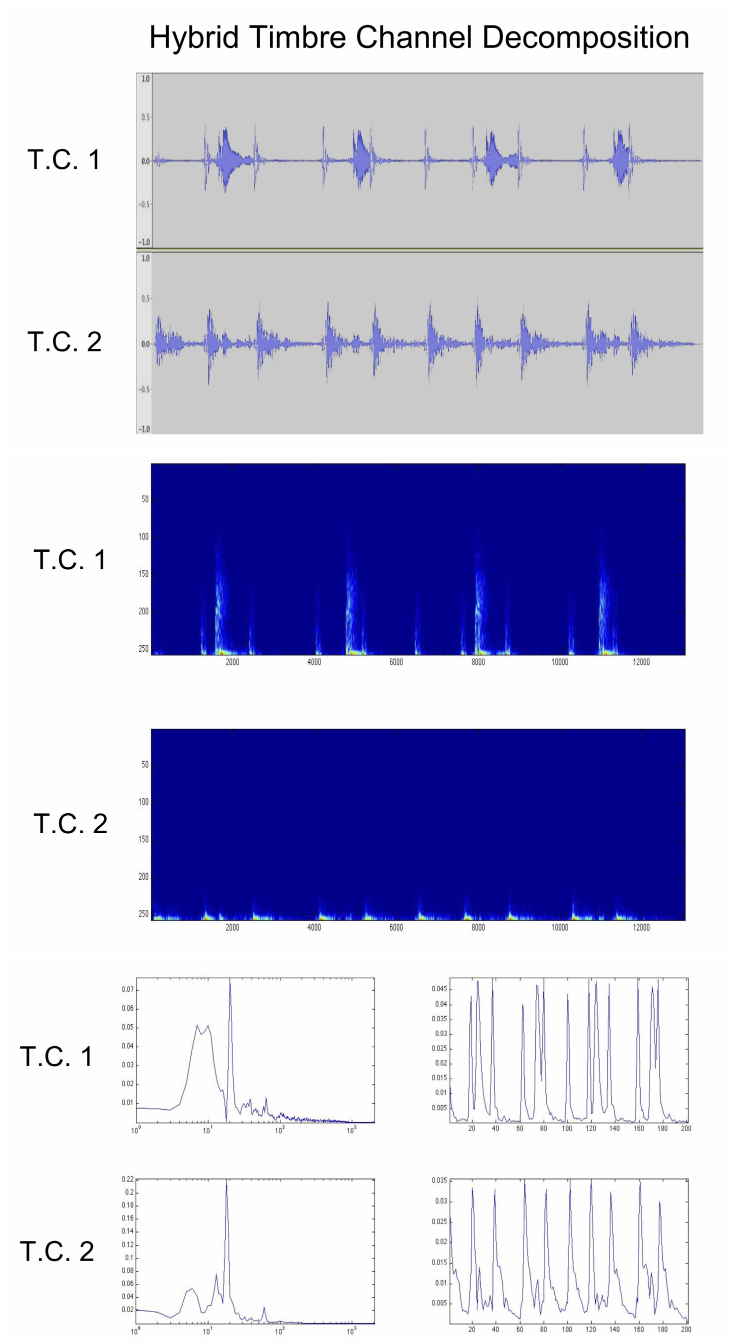


Figure 3.6: Hybrid Timbres: where the percussive harmonic (with added noise) and rhythmic percussive content is blurred, creating hybrid mixtures between each timbre. The top audio track contains more drum timbres, while the bottom audio track contains more harmonic timbres, shown in three representations: resynthesized waveforms,log-spectrograms, and PLCA H, W components.

3.4 A Digital Musicology Approach to Latent Structure

Digital musicology was a term coined by Roland and Downie (2007) to describe computer analysis musicology. This section explores digital musicology using the latent structure analysis of two compositions by twentieth-century composers John Cage and Gérard Grisey. The pieces were selected based on the following criteria: (a) They must have multiple prior analyses, which allows for a comparison of techniques and claims; (b) there must be recorded acoustic audio because it is more difficult and more realistic; and (c) there must be contrast between the pieces, which is necessary for demonstrating adaptability.

3.4.1 John Cage's *Fourth Interlude for Prepared Piano*

The *Fourth Interlude* belongs to the series *Sonatas and Interludes* (19461948) and was written by Cage during the Second World War. It was in 1945 when Cage met the Indian composer and tabla player Gita Sarabhai, who studied with Cage and subsequently introduced him to Indian music and culture. Of equal importance, particularly with respect to the *Fourth Interlude*, is Cages contact with Javanese Gamelan music through the lectures of Henry Cowell (Ingram 2006).

It was through a re-imagining of the piano as these exotic instruments that Cage arrived at the particular arrangement of screws, rubber, and plastic (shown in Figure 3.7):

> I placed objects on the strings, deciding their position according to the sounds that resulted. Having those preparations of the piano and playing with them on the keyboard in an improvisatory way, I found melodies and combinations of sounds that worked with the given structure. Just as you go along the beach and pick up pretty shells

Figure 3.7: A Steinway Concert "D" Grand Piano with the prepared setup for John Cage's Sonatas and Interludes Cage (1940-47).

> that please you, I go into the piano and find sounds I like, (Cage 1995).

What are these sounds, and how many distinct sounds is Cage talking about? Interpretation of sound, in this case, refers to the idea of timbre because an unprepared piano has an unusually normalized timbre across the entire range of the instrument, and preparing it would change this attribute. To answer the question concerning the taxonomy of sounds in the Sonatas and Interludes, latent structure analysis was performed on the Fourth Interlude using PLCA2d because it is noticeably rich in timbral content (Perry 2005).

The analysis method consisted of a 20-component decomposition of the first 30 seconds of the *Fourth Interlude* performed by Boris Berman (Cage et al. 2005). To perform this analysis, the audio is converted from stereo to mono, followed

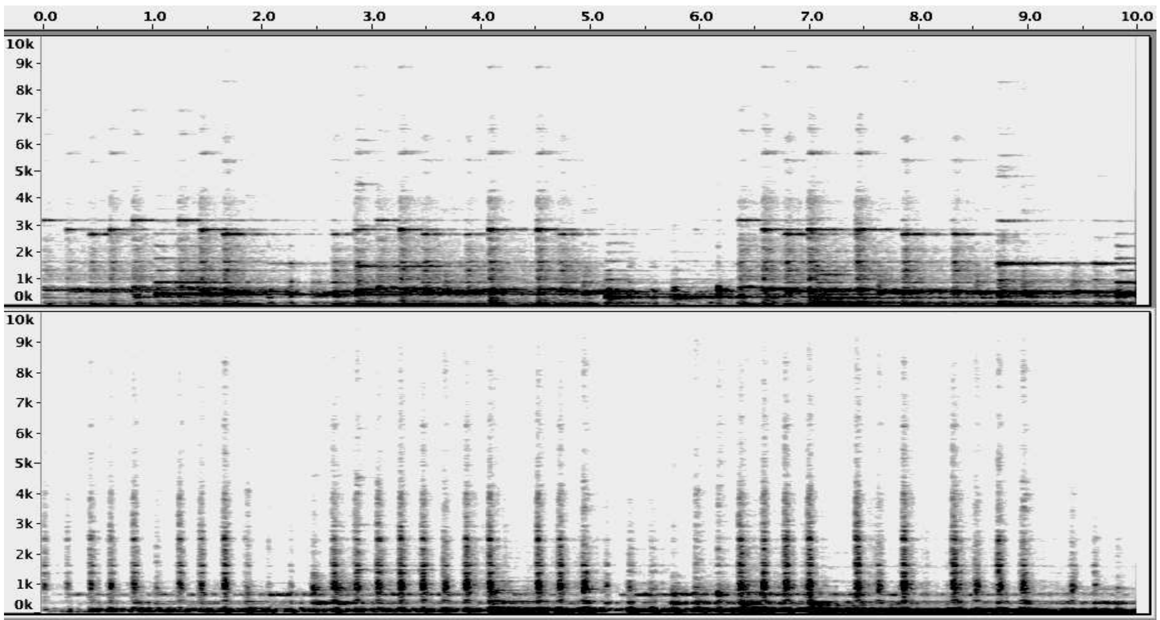


Figure 3.8: The spectral decomposition of the first 10 seconds of John Cages *Fourth Variation for Prepared Piano*. The top spectrogram shows the harmonic timbres, and the bottom spectrogram shows the percussive timbres.

by a transformation to the time-frequency domain using STFT with a 2048 sample window and a 1024 sample hop. Using the PLCA2d algorithm, components were extracted using overlap-add re-synthesis (George and Smith 1992), with the code listed in Appendix B.0.1.

With the components resynthesized, they were auditioned and sorted according to timbre and produced only two clusters. Figure 3.8 shows two spectrograms of the resynthesized components, with the top spectrogram consisting of harmonic-pitch information and the bottom consisting of percussive-noise information. In essence, this analysis suggests there are only two main timbres, which is remarkable considering the different material Cage used to prepare the piano.

Another possibility is that for this specific piece the interaction between the timbres created a situation in which the underlying design of the piano became exaggerated, which means the screws accentuated the hammer striking the string, and the rubber and plastic dulled the effect of the hammer. When the parts interact very quickly in this music, there is an ASA stream-like effect with regard to percussive and harmonic timbres. For this example, at least, the seemingly complex timbres separate into relatively simple parts, which when combined create an intricate tapestry of sounds.

3.4.2 Gérard Grisey's *Partiels Pour 16 Ou 18 Musiciens*

In a more recent composition by Gérard Grisey, latent structure is examined by analyzing spectral music. In this case, there is an explicit reference concerning the source material for the composition (Arrell 2002), which consisted of an E2 played on a concert trombone. This allows for the direct comparison of what the composer notates in the score to the content of the trombone note.

For this analysis, the method is similar to the Cage analysis, but eight components were selected using trial and error. Instead of a short excerpt of the piece, however, the entire first section (i.e., the first 4 minutes) of the composition was analyzed.

Using the code from B.0.2, the audio recording of *Partiels* (Grisey, 1981) was converted from stereo to mono. Then it was transformed into the time-

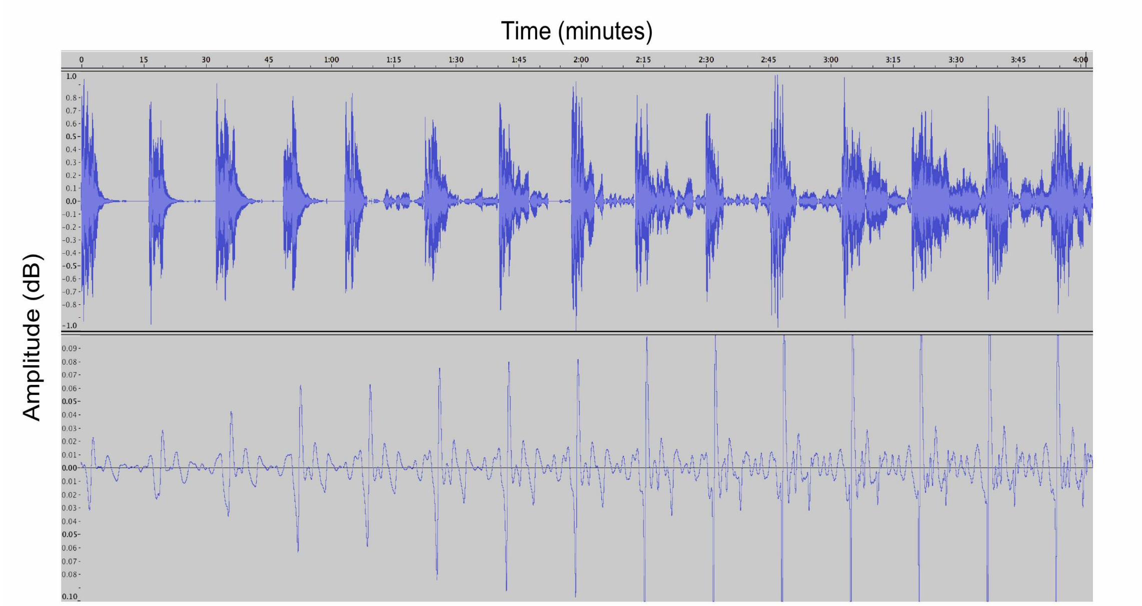


Figure 3.9: A comparison between an E2 Trombone note attack elongated via time stretching (below), and the first component from a eight-component PLCA spectral decomposition, containing similar timbres consisting of string bass, low brass, which repeatedly articulate the fundamental (above). The striking connect between the two waveforms is the congruent periodicities of the amplitudes.

frequency domain using the STFT function with a 2048 sample window and 1024 sample hop. Using the PLCA2d algorithm, components were extracted using overlap-add re-synthesis (George and Smith 1992).

As expected, the frequency content matched closely with the spectral information in the trombone sample. Surprisingly, a closer examination of the first component containing the icon-like double-bass fundamental with the trombone sample shows the structure of the first section was somehow tied to the structure of the trombone sample, with amplitude peaks in the trombone sample. Using a basic re-sampling of the trombone sample, it was possible to dis-

Figure 3.10: ”Partiels” by G. Grisey, rehearsal number 12, page 14 (Grisey 1975).

cover the right phase pattern for two tracks. This suggests that Grisey used the amplitude information as well as the spectral information to inform structure in the music. Taking the periodic structure analysis described above further, the remainder of the components were examined, and it was noticed that Grisey used these periodic waves to emphasize different parts of the original spectral information over time, shown by the solid boxed areas, and gradually bring everything into phase. The dotted-line boxes show the most dramatic of these offsets, which occur between the first and sixth components. This continues until every single component eventually comes into phase at rehearsal letter 12, shown in Figure 3.10. Analysis of the score also suggests that Grisey wanted very exact

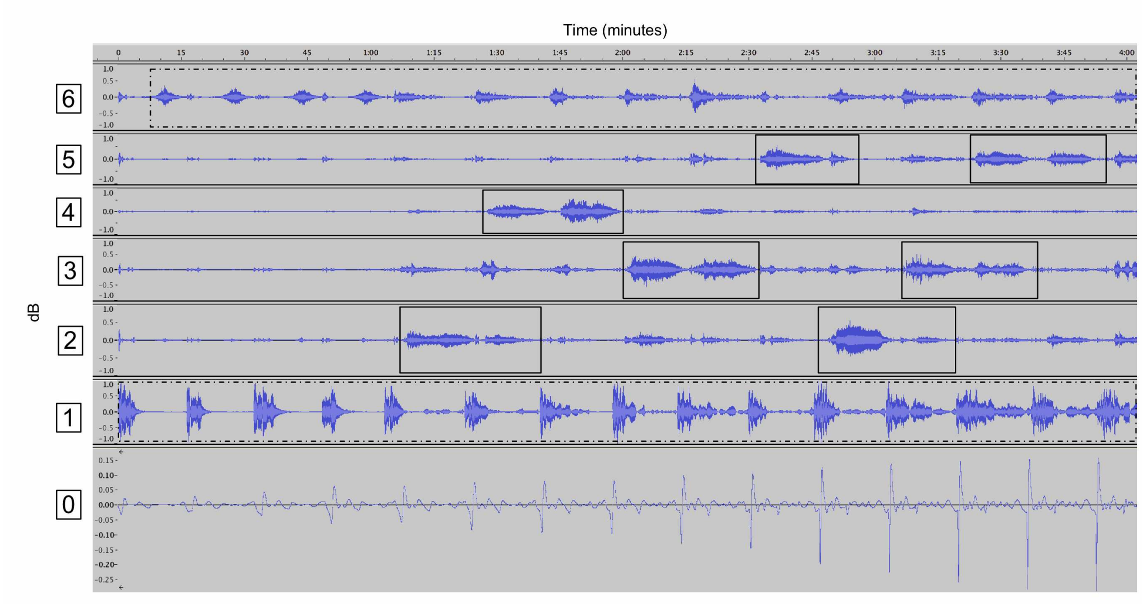


Figure 3.11: An expanded analysis of Figure 3.9. Shown here are six of the eight PLCA extracted components. The null track is the E2 trombone, while the first component is the articulated fundamental. The remaining components contain partials from the original trombone analysis used by Grisey.

timings for events, with seconds and finite duration markings on every page. This makes for a compelling argument that Grisey made every attempt to structure the music around the source material, and with latent component analysis, it is possible to see the detail in the underlying periodic mechanisms correlated to the audio source material.

3.5 Summary

This chapter suggests music audio is a mixture of interdependent events. When music events are sufficiently independent, components can be extracted that reflect distinctive sound and music objects. The methods and the historical precedent for these techniques were discussed, with particular examples in spectral music, and the theories of Pierre Schaeffer concerning sound objects, music objects, and magnets in music audio (Schaeffer 1967).

The latter half of the chapter examined the primary algorithm used in this dissertation (i.e., PLCA) with respect to digital musicology, and latent structure analyses were conducted on pieces by John Cage and Gérard Grisey.

In both cases, latent structure analysis using the PLCA2d algorithm produced new ways of interpreting the musical content and design. Specifically, the Cage decomposition raises questions about how timbre operates in the *Sonatas and Interludes for Prepared Piano*. The Grisey analysis demonstrates that deeper connections exist between the periodic structure of the music and the trombone audio signal cited as a source material for the piece.

CHAPTER 4

GROOVE RETRIEVAL IN MUSIC AUDIO DATABASES

Latent structure is now examined in a macro-analysis using a groove retrieval task. Consider first the following characterization: Western popular music often has a repetitive or quasi-repetitive percussive background called a groove. The instruments or sounds assigned to this role typically are quite limited and include a bass or kick drum sound with a low center frequency, a center mid-frequency noise burst that can be an acoustic or synthesized snare drum, and a high-frequency bright high-hat cymbal that is also either acoustic or synthesized. There is, of course, great variation in these basic sounds, yet it is argued here that *timbre channels* are a phenomenon that forms as a result of consistent information. In other words, while the music may change considerably, the instruments in the percussive backgrounds are usually static and consistent in their spectral template. This suggests that these sounds occupy three special cases of timbre channels that occupy three general regions in the average human frequency spectrum (i.e., 25—19,000 Hz). Given these assumptions, it is hypothesized that rhythm patterns are contextualized by timbre, and similarity in real-world rhythm depends on some generalized notion of timbre.

4.1 Basic Properties of Rhythm in the Signal Domain

Knowledge about signal-level transforms is essential for understanding rhythm in music audio because the components, like those discussed in chapter 3, are extracted from features generated by these methods. Common transforms include the following: (a) FFT, (b) STFT, and (c) Center-Q Fourier ransform

(CQFT). Specifically, understanding the effects of the transform parameters were critical to the rhythm experiments presented in this chapter.

STFT is a standard method for extracting rhythmic information from a musical signal. Unlike FFT (Equation 4.1), which does not encode change in frequency over time, the STFT rectifies this issue by employing a time-limiting windowed version of FFT (see 4.2). Modifying FFT to include windowing means that the STFT can be used to encode temporal events. The FFT can be expressed in the following way:

Fast Fourier Transform (FFT)

$$\hat{s}(\omega) = \int_{-\infty}^{\infty} s(\tau) \cdot e^{-i\omega\tau} d\tau \tag{4.1}$$

where $\hat{s}(\omega)$ is the Fourier Transform belonging to each continuous frequency ω, $i = \sqrt{-1}$ and is known as the imaginary identity with the natural exponential base e. This equation is modified to include frame segmentation by including $\overline{h}(\tau - t)$, s. th. $\overline{h}(t)$ is the complex conjugate of the window function. This inclusion in the FFT product generates the following equation known as the Short-Time Fourier Transform (STFT):

$$\hat{s}(t, \omega) = \int_{-\infty}^{\infty} s(\tau) \cdot \overline{h}(\tau - t) \cdot e^{-i\omega\tau} d\tau \tag{4.2}$$

In STFT, the $h(t)$ time scale, unlike the continuous time scale in the FFT, is independent of the harmonic number $f = \tau$, and the window function is the same for all harmonic components. This means that when the STFT is assumed to be discrete it is also assumed that all harmonic components will be in ratio with $h(t)$. The net effect of this assumption is that blurring will occur if the frequencies for a given window do not match the harmonic ratio within the time scale $h(t)$. To mitigate the effects of blurring, a probability density envelope is applied over the basis function and can be selected based on the preference of the user. Typical functions of this type include the Gaussian window, Hamming window, or Blackman-Hamming windows, respectively. The non-continuous block version of the STFT, shown in Equation 4.3, is the form used in digital signal processing (DSP) STFT algorithms:

$$F_{\infty}(m,k) = \sum_{n=-\infty}^{\infty} s(n) \cdot h(mR_A - n) \cdot e^{\frac{-j2\pi nk}{N}} \tag{4.3}$$

This function can be modified to compute the root mean square average (RMS-Average), or magnitude, over all consecutive frames of the STFT (see equation 4.4):

$$S_M = \sqrt{\frac{S_m^2}{s}} \tag{4.4}$$

For each block m of the STFT, the RMS-Average is computed for all s bins and results in the vector S with M number of frame samples, and the RMS-Average over all frequency bins captures the percussiveness of the mixture for M bins. This method was first introduced by Scheirer (1996) in order to analyze meter and rhythm. Figure 4.1 illustrates the difference between rhythm articulated as waveform amplitude and RMS-Average across STFT frames. It also illustrates that rhythm content is observable from both the waveform and the STFT-RMS envelope plots by changes in energy over time. The STFT-RMS simplifies the

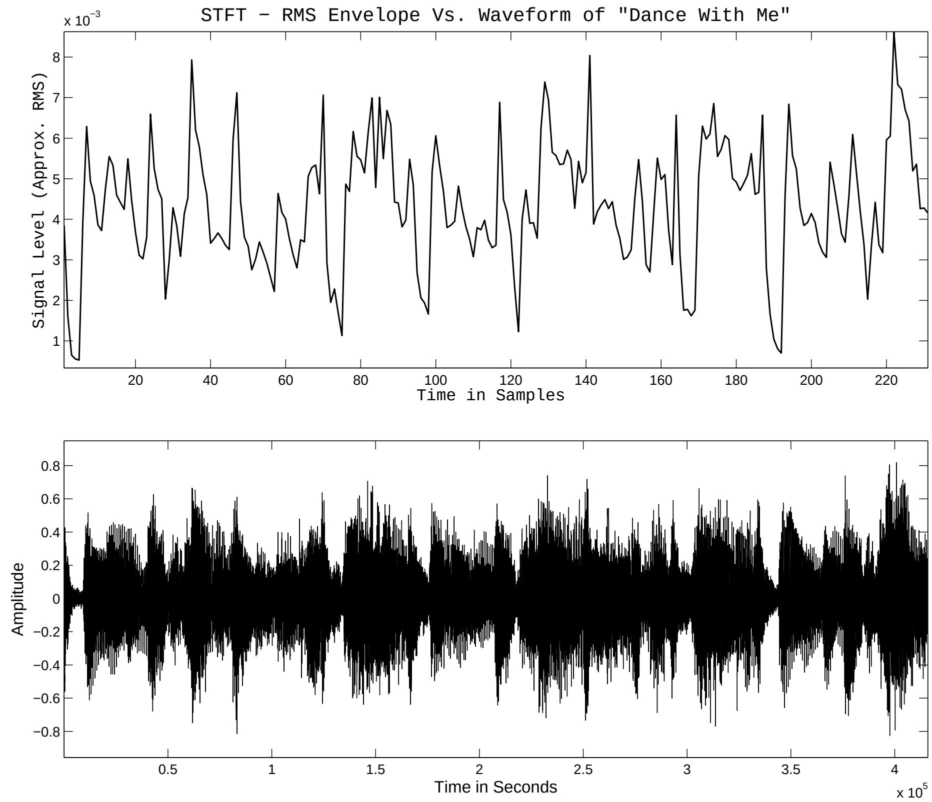


Figure 4.1: The opening of Dance With Me showing the rhythm represented as waveform amplitude (below) and STFT-RMS envelope across STFT frames (above).

rhythmic content of the waveform by representing a magnitude-only form of the rhythmic information.

What should also be apparent is the lack of information required to identify the sources that created these rhythmic events. These representations are mixtures in which the information identifying how different sources contribute to percussive events is unknown. To solve the analysis for mixtures of sources in audio, source separation techniques can be employed to determine patterns in frequency-time and envelope-time spaces to produce independent components. It is the aim of latent rhythm stream analysis to use source-separated components as representations of elements belonging to sub-mixtures.

4.1.1 Rhythmic Information in Source Separated Components

Interest in spectral decomposition exploded in music information research over the past 10 years, with recent, notable examples including GAP-NMF (Hoffman et al. 2010), and IS-PLCA (Weiss et al. 2010). The previous application of source-separation techniques addressing rhythm in MIR mainly centered around audio classification and beat tracking (Tsunoo et al. 2009), tempo estimation (Chordia and Rae 2009, Woodruff et al. 2006), and most predominantly in drum and music transcription applications (Gillet and Richard 2008, Plumbley et al. 2007, **?**), and rhythm analysis (Barry et al. 2005, Orife et al. 2001).Source separation is used in these applications to un-mix sources and improve information retrieval or processing, an assumption shared by the groove retrieval system discussed here.

For all of the retrieval experiments mentioned above, the stages of processing require reconstruction of components into audio, with the exception of IS-PLCA, which directly analyzes components to perform segmentation estimation. Groove retrieval using timbre channels differs from these other systems because analysis and similarity measurements come directly from the components. Using source separation for the purpose of un-mixing sources appears to be based on the assumption that performing analysis on audio reconstructed from components is equivalent to performing analysis on non-source-separated audio.

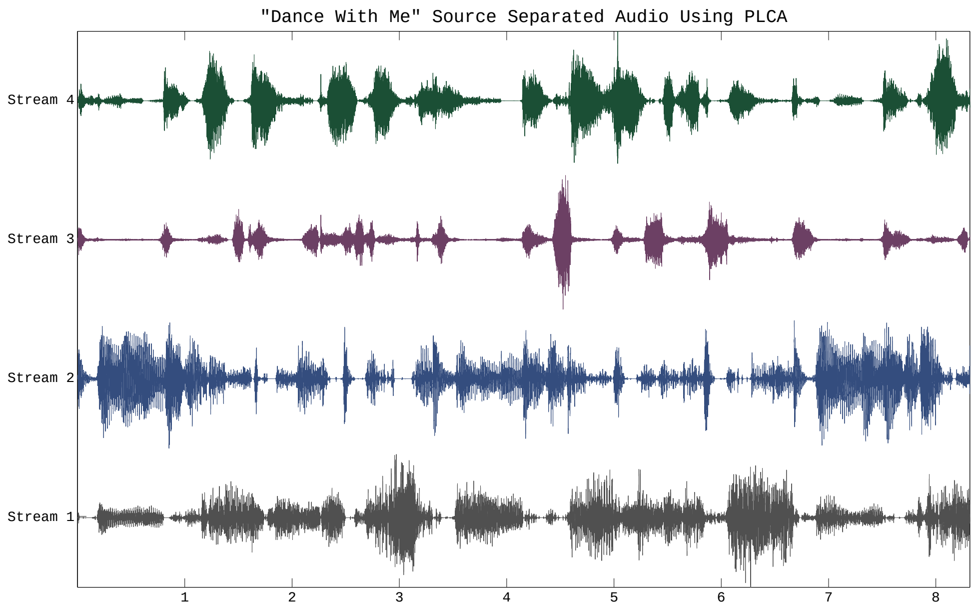


Figure 4.2: A plot of four reconstructed audio segments extracted from the song Dance With Me using PLCA. The x-axis is time in seconds, and the y-axis is amplitude.

Figure 4.2 illustrates reconstructed audio using PLCA components extracted from the previous musical example, Dance With Me. Convolving these reconstructed audio segments results in a mixture that is equivalent or identical to the original mixture. This is the same way an inverse FFT or inverse STFT will produce the original audio processed by the FFT and STFT transforms.

Compare the reconstructed audio representation in Figure 4.2 to the non-reconstructed temporal components shown in Figure 4.3. The main difference between the two is that the frequency components are left out and only the change in magnitude over time remains. It is equivalent, in a sense, to the RMS-STFT rhythmic content shown in Figure 4.1, except Figure 4.1 represents a mixture, while Figure 4.2 is a set of elements belonging to that mixture. The mixture

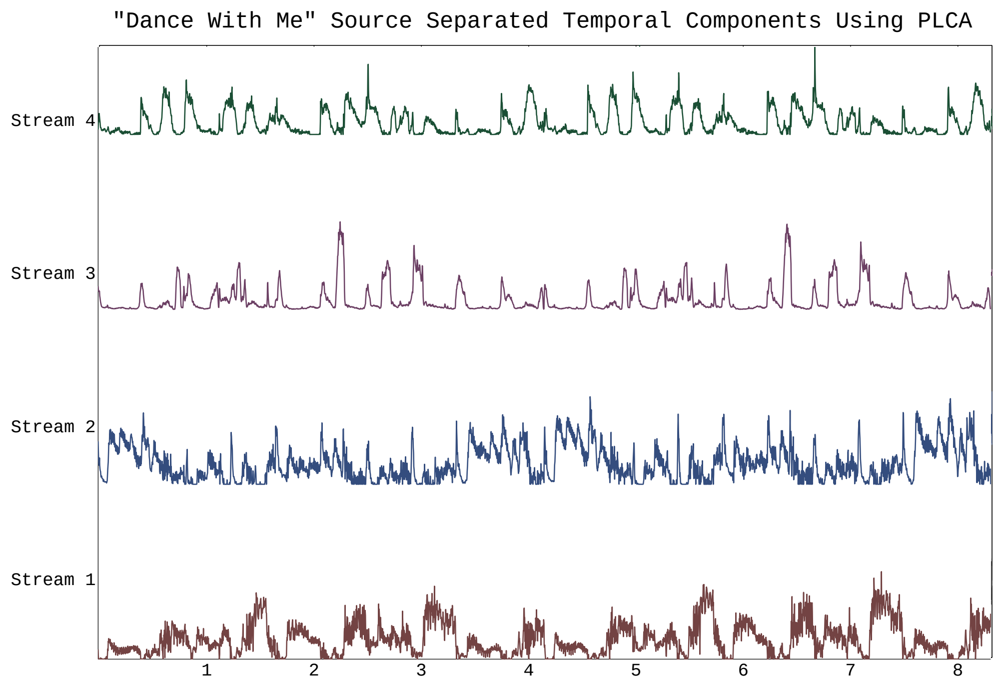


Figure 4.3: A plot of temporal components extracted from the song "Dance With Me" using Probabilistic Latent Component Analysis (PLCA), where the x-axis is time in seconds, and the y-axis is magnitude.

paradigm forms the foundation of groove retrieval (see Section 4.2.3).

Probability Distributions of Rhythmic Envelopes

Figure 4.4 illustrates the sub-plots of source-separated components extracted from an audio sample of the Amen Break. The left column shows the frequency *w* marginals, while the right column shows the temporal envelope *h* marginals. The rows correspond to bass drum, high-hat cymbal, and snare. Each component set of *w* and *h* components is distinctive from the other set, and the content of each *h* component is fairly consistent to a particular envelope shape. Therefore, it can be hypothesized that if there is a correlation between components from different songs with similar shapes then there will be correlations between rhythmic patterns.

4.2 ISHKUR Groove Retrieval Experiment

Information retrieval experiments are a standard way to measure the performance of MIR models because they require the identification of relevant items and the relative similarity of these items. This identification is made using distance metrics (e.g., Euclidean distance) (Foote et al. 1997) or by clustering extracted features, such as MFCCs (Casey et al. 2008).

The central objective of the retrieval tasks described in this section was to retrieve cross-timbral rhythm patterns extant in sub-mixtures, that is, a groove. Grooves can be thought of as the percussive-rhythmic content in a song that provides a reoccurring or semi-reoccurring background that ties together the musical material on the surface. These sub-mixtures can be remixed into songs in a variety of ways, ranging from audio transformations to complete re-sequencing.

Another important aspect of grooves is their persistent reuse in different songs across different genres. DJs, song makers, and producers often use samples of grooves, or break beats that are popular in dance genres, to reference specific styles and artists. Among the most famous and widely-used samples are the "Funky Drummer" (Brown 1969), "Amen Break" (Winstons 1969), "Big Beat" (Bang 1989), and "Anthem Breaks" (Veal 2007).

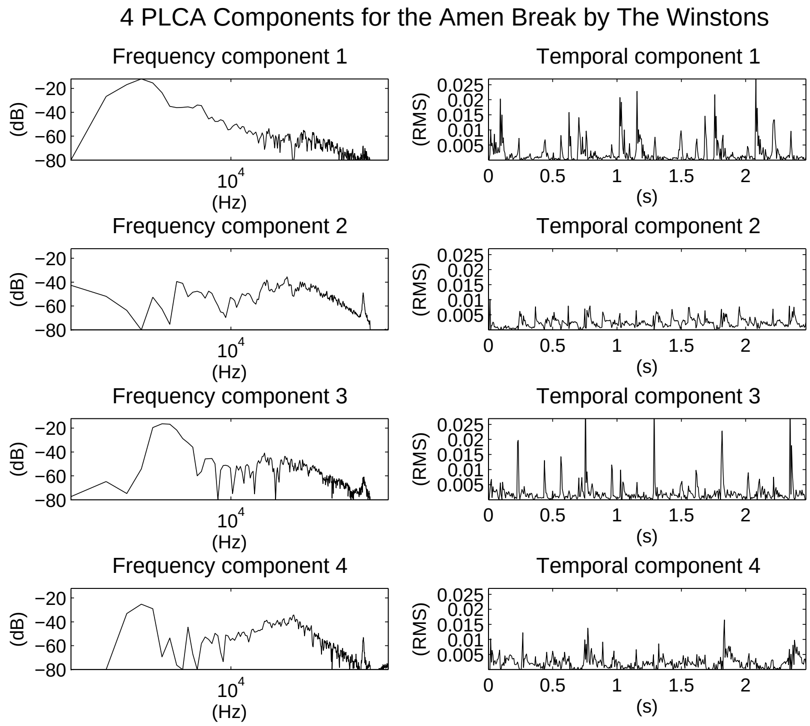


Figure 4.4: The source separated components from PLCA of the Amen Break. The frequency components (left column) have sub-plots consisting of spectral features, with the x-axis in frequency (Hz) and the y-axis in dB. The temporal components (right column) have sub-plots consisting of temporal envelopes in which the x- axis is time in seconds, and the y-axis is the RMS values corresponding to the signal magnitude of similar sources. However, as stated at the beginning of this chapter, doing so proved to be unsuccessful, either by classifying the w components, or by classifying the MFCC features derived from the components themselves. Instead, a Global-PLCA method was proposed in Qingyuan et al. (2011) aimed at extracting components that shared features consistent across an entire collection of songs. It was this method that ultimately beat the precision/recall baseline consisting of sub-band decomposition (see Section 4.2.5).

.

To summarize, the cross-genre aspect of grooves combined with a variety of transformations of the original samples makes groove retrieval a sufficiently hard similarity problem, but it lends itself to a latent structure approach. The sections that follow outline the methods and results of groove retrieval performed on the 1138 song ISHKUR dataset.

4.2.1 Overview of Bregman Groove Retrieval Implementation

The groove retrieval experiments were first written in Matlab, but following subsequent revisions, it was ultimately decided that a new implementation in Python was necessary as a result of the increasing complexity of the data structures. This complexity was, in part, a result of the fundamental aspects of the data, including track length and sampling rate. The system itself also increased the complexity of the system. For every window from a song segment, there were three components (i.e.,$\mathbf{w}, \mathbf{h}, \mathbf{z}$) for every ρ extracted sources. This means there are 3 components per frame of analysis, creating a $\rho!$ combinatorial relationship between the number of sources and the possible ways the sources could correlate to other sources.

These factors precipitated the need for a more sophisticated source separation method and a faster, more robust analysis system. This need culminated in two important contributions: (a) Bregman MIR library by Michael Casey and (b) the Global-PLCA algorithm (Qingyuan et al. 2011) presented in section 4.2.3. The process of managing MIR experiments is complex, and the Bregman library provides experimental stages that interface with the AudioDB C framework and is open and accessible as an online Python library that can be downloaded (Casey). Three general stages of processing are required to run the following experiments:

- **Generate keys**: For a given set of audio files, generate unique keys, and if possible, correspond keys with ground-truth information.
- **Extract features**:A specified chain of pre-processing and processing on audio files, which can range from one feature to *n* features.
- **Evaluate results**: Using a range of different methods, including distance measures, clustering, etc. these associated functions generate a set of results that can then be analyzed using standard information retrieval measures.

Another advantage of using AudioDB is the built-in record-keeping system that enables all features to be automatically saved as sessions. This allows the researcher to quickly move forward without the risk of losing data. For these reasons, this library was selected as the superior choice for evaluating the groove retrieval experiment.

4.2.2 Method

The ISHKUR Data Set

A dataset was prepared consisting of 1138 song excerpts from the Ishkur Electronic Music Guide (Doe 2010). The excerpts range from 3 seconds to 90 seconds. This particular dataset was selected according to the following criteria:

- Audio tracks containing identifiable rhythmic beats or grooves.
- A wide range of stylistic content across many genres of music.
- Audio quality reflective of online databases.
- Accessible to researchers as a free online download.

Ground-Truth Collection Methods

The ground-truth requirements for this experiment involved having listeners identify grooves from the ISHKUR dataset. These similarity listening lists were compiled by two expert musicians: (a) the author, who has more than 15 years of composition and performance experience (from this point on identified as Listener I) and (b) Amir Sa'id, a professional DJ from New York City named (identified as Listener II). Both listeners were asked to independently evaluate all the songs in the database and decide on a personal strategy for sorting the song excerpts into groups based on similarity of groove. For simplicity, groove similarity was restricted to one selection per track. In situations where listeners encountered more than one groove in a given song excerpt–either in sequence or as a mixture–they were asked to select the most obvious or predominant groove.

Figure 4.5 shows the interface created in the interactive visual programming environment Max 5 each listener used to sort the categories. One category, the non-beat category, was provided in order to identify tracks that contained no percussive material from relevant tracks. While this was not set up to be a perceptual experiment, the listeners choose their own categories in order to mitigate the effect of forced-choice selection of categories because categorical similarity can be greatly influenced by arbitrary constraints (Marcel 1983).

Figure 4.5: The evaluation interface used by expert listeners to sort the ISHKUR dataset into categories by groove.

The two expert listeners completed the dataset markup in 3 weeks, and in that time, they did not discuss the category assignments or any particular details about the content of the song segments. The genre categories were not available either, with the rationale that prior categorizations might influence the track category assignments. As mentioned earlier, the only provided category was non-beat. Listener II did not follow this instruction, however, and instead used the non-beat category to note anything, percussive or otherwise, that was not relevant. Aside from this detail, using a basic set intersection between the two ground-truth markups, four shared categories were identified between Listener I and Listener II and subsequently identified as the areas of interest in the following retrieval experiments.

Data Set Content and Organization

The dataset was compiled from a freely available compressed file containing examples from the *Ishkur Electronic Music Guide* (Doe, 2010). The clips range between > 3 seconds and < 40 seconds, with the lowest sampling rates at 11,050 Hz and the highest at 44,100 Hz. All clips were treated with pre-processing using Audacity (2011), which converted the stereo tracks to mono, audio levels were normalized, and sampling rates were converted to 44,100 Hz.

All audio clips were organized into 183 folders in the dataset, corresponding to the genre categories on the *Ishkur Electronic Music Guide* website. These categories ultimately proved impossible to work with because some genre files contained only one clip, while other files contained up to 10 clips. Instead, an evaluation tool was devised to help the expert listeners determine groove categories without prior knowledge of these genres.

Category Attributes from Expert Listeners

The two listeners did not have access to the genre information while marking; therefore, they were asked to provide their own interpretations of categories based on groove similarity (i.e., with the simple question "Which of these grooves are similar to one another?"). The following tables show the self-described categories by both listeners and the number of song elements in each category, ranked from highest number to lowest number of song excerpts per category.

Each listener was further asked to provide a description of the groove content observed for each category identified in the sorting task. It is interesting

Table 4.1: Listener I Self-Assigned Groove Categories

Groove Category	Number of Songs
Disco and Funk Beats	254
Techno	218
Break-Pulse mixture	106
Erratic/Noise	90
Drum Chatter	75
Repeating Bass Pulse	71
Non-Beat	58
Swing Breaks	57
Two-Step	57
Straight Break	44
Amerifro Drumming	40
Non-Break Trance Beats	28
Triplet-Polyrhythmic	15
Short-Short-Long	15

to note that Listener I tended toward classifications based on rhythm similarity, while Listener II based similarity as much or more on timbre than rhythm content. A complete list of these descriptions can be found in Appendix C.

Figure 4.6 shows the proportion of songs per category over the whole dataset selected by Listener I, while Figure 4.7 show the proportions for Listener II. It is interesting to note that Listener IIs largest category of songs was the not relevant (labeled as non-beat) category, and the largest proportion of songs across the relevant categories in both listeners markup were more general than the cat-

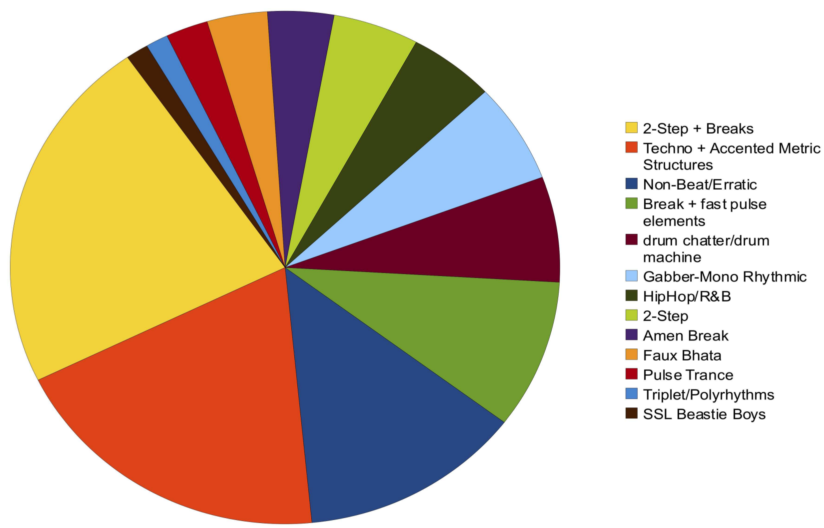


Figure 4.6: Listener-identified groove categories selected by Listener I.

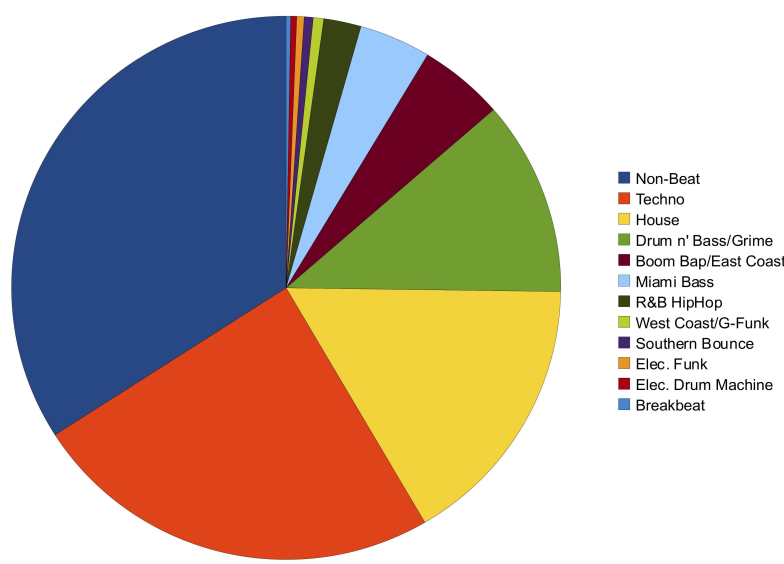


Figure 4.7: Listener-identified groove categories selected by Listener II.

Table 4.2: Listener II Self-Assigned Groove Categories

Groove Category	**Number of Songs**
Non-Beat	388
Techno Beats	277
House	186
Drum n' Bass/Grime	132
Boom Bap/East Coast	57
Miami Bass	48
Rhythm n' Bass/ HipHop	25
West Coast/G-Funk	7
Southern Bounce	6
Elec. Funk	5
Elec. Drum Machine	4
Breakbeat	3

egories with fewer song excerpts. This suggests one of two possibilities: (a) some grooves are more generic than other grooves or (b) certain genres greatly influence the specificity of classification. It is worth noting that it is most likely an amalgamation of the two effects, at least in the case of this dataset markup.

In order to pursue a testable hypothesis with regard to groove identification, two different strategies were taken (see section 4.2.5). The first strategy was to test each of the listeners categories individually, and the second strategy was to take the set intersection between both markups, select the best matching intersections between both listeners, and consider these categories the similarity

Table 4.3: Top Three Categories in Set Intersection between Listeners

Listener I Category	textbfListener II Category	**Number of Song Intersections**
Techno	Techno Beats	141
FunkDiscoHouse	House	114
Drum Chatter	Drum n' Bass/Grime	50

space for the experiment. Table 4.2.2 shows the top three categories that intersect between the two expert listeners. Notice genre or sub-genre labels. This might suggest that certain genres have more consistent grooves across many different songs, or some music genres are better defined by their groove content than by other musical features.

To summarize, two listeners were asked to sort song excerpts into categories of their choosing. These categories were expected to reflect the groove content of the tracks. In addition to providing a categorical choice for each excerpt, they were asked to give each category a name and description. Three categories had the highest set intersection between the two listeners: (a) Techno, (b) House, and (b) Drum n Bass.

Baseline Criterion

The Scheirer (1996) sub-band decomposition via comb-filter technique was used to extract the pulse or tactus. Constant-Q Fourier Transform and Constant-Q Fourier Transform PLCA separated outputs were compared in a series of experiments on a synthetic dataset consisting of randomly generated beat patterns synthesized into three rhythm streams with timbre characteristics that emulated a high-hat cymbal, snare, and kick drum. When mixed together, they formed the equivalent of a synthetic break-beat groove. Each of the 100 patterns was then called ground-truth patterns that corresponded to a set of variant patterns reflecting random incremental changes to the original patterns (i.e., one beat changed, then two).

4.2.3 Retrieval by Groove: Application of Timbre Channels to Rhythm Analysis

Retrieval by groove generally describes a process for extracting sub-mixtures consisting of grooves belonging to music tracks and comparing them using Euclidean distance metrics, a common method for measuring similarity. The basic procedure of the model is explained below.

Pre-Processing

The ISHKUR dataset was preprocessed using unix batch scripts in which each mp3 track was converted to a .wav file and converted to a 44,100 Hz sampling rate for consistency across all tracks because mp3 sampling rates vary between 11025 Hz and 44100 Hz. PLCA computes component extraction on only mono audio files; therefore, all audio was converted to mono and normalized to account for possible album effect (Downie 2008).

An STFT was then performed with a 92 ms window and 50 ms hop, followed by a re-binning stage consisting of 12 constant-Q frequency bands per-

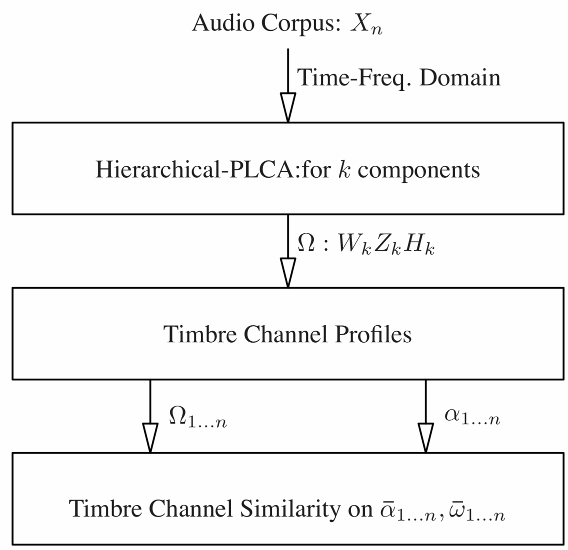


Figure 4.8: Overview of the Bregman groove retrieval algorithm: Hierarchical-PLCA extracts consistent with global timbre channels using a multistage PLCA decomposition. The timbre channels are evaluated and given a probability weight , which is used to weight similarity in the final Euclidean or Bhattacharyya-Euclidean distances between the target track and all other songs in the dataset.

octave. Then a log magnitude and inverse discrete cosine transform was performed (Abdallah and Plumbley 2004) by applying a lifter that yielded 10 cepstral coefficients starting at the second. The features were then inverted back to a linear amplitude constant-Q spectrum to make the features separable for PLCA analysis (Qingyuan et al. 2011).

Rhythm Extraction Process

The problem of correlating global components was encountered early in this research. Take a simple example: Audio track A has many instrumental sources (i.e., orchestra, vocals, and drums), and track B has only vocals, guitar, and drums. Even with the correct component estimation, which is highly unlikely in an unsupervised context, how does one know which tracks correlate to one another?

In the initial design of the groove retrieval system, unsupervised and supervised clustering algorithms were used to find a correspondence between PLCA components. However, the use of clustering in this process proved unwieldy as a result of propagating errors from inaccurate clustering, possibly resulting from the wide variability of the content in each component. Moreover, PLCA was used with fixed-rank decomposition parameters, which again might have contributed to the propagation of errors from clustering by decreasing any consistency between the PLCA components.

Hierarchical Probabilistic Latent Component Analysis (H-PLCA)

A new method called Hierarchical-PLCA (H-PLCA) was proposed to reconcile the correspondence problem in the latent component extraction of an audio dataset (Qingyuan et al. 2011). This method used a three-level hierarchical iterative analysis requiring two types of priors: (a) an entropic prior and (b) a Dirichlet prior. The model uses the entropic prior to measure the contribution of components at each of the three hierarchical levels. The Dirichlet prior is then employed at the final stage of extraction to adapt the local features in each track to the global basis features. The following describe the stages of H-PLCA:

- **Stage I Local *W*-Step:** Per-Track PLCA Decomposition
- **Stage II Global *W*-Step:** Universal Latent Spectral decomposition on all *W* Components across all tracks

- **Stage III Local *H*-Step:** Per-Track Amplitude-Time, H, component extraction using the universal basis distributions

The goal of this implementation is to extract consistent, globally contextualized timbre channels (Qingyuan et al. 2011). H-PLCA with timbre channels solves three important problems with previous methods: (a) the rank estimation problem for global component extraction, (b) finding correspondence between components, and (c) removing the requirement to train data for clustering. In essence, this method provides a way to access timbre channels contextualized by the surrounding data, or what is sometimes referred to as a context similarity measure (Jeh and Widom 2002). Timbre channels in this context emulate aspects of streaming from Bregman's auditory objects described in chapter 2. Similar to auditory streams, timbre channels extracted from a song database present consistent and repetitive spectral information that enables a listener to make reliable associations between one stream and another stream.

4.2.4 Timbre-Rhythm Similarity

To compute rhythm similarity, the rhythm distance method was used (Qingyuan et al. 2011). This method can be defined in a probabilistic framework as the Bhattacharyya-Euclidean distance function between two tracks, i and j, where the Euclidean distance between two latent time-amplitude distributions ($\hat{P}^{(i)}(t|l)$, $\hat{P}^{(j)}(t|l)$) are calculated, weighted by the prior probabilities of each time-amplitude distribution ($\hat{P}^{(i)}(l)\hat{P}^{(j)}(l)$):

method	prec $1-3$	prec $1-5$	prec $1-10$
Subband d_{Avg}	0.45	0.31	0.19
Subband d_{Rhythm}	0.24	0.23	0.17
H-PLCA d_{Avg}	0.48	0.34	0.22
H-PLCA d_{Rhythm}	0.50	0.36	0.23

Table 4.4: Groove retrieval for sub-band and H-PLCA amplitude-time features using Euclidean average distance and the Joint Timbre-Rhythm Similarity Measure.

$$d_{Rhythm}(i,j) = 1 - \frac{\sum_{l=1}^{l=L} \hat{P}^{(i)}(l)\hat{P}^{(j)}(l) \sum_{l=1}^{l=L} \left\{ \hat{P}^{(i)}(l)\hat{P}^{(j)}(l) - d_{Euc}(\hat{P}^{(i)}(t|l), \hat{P}^{(j)}(t|l)) \right\}}{\sum_{l=1}^{l=L} \exp(\hat{P}^{(i)}(l)\hat{P}^{(j)}(l))} \tag{4.5}$$

The canonical form of the Bhattacharyya metric (Kailath 1967) $BC(P,Q) = \sum_{x \in X} \sqrt{p(x)q(x)}$ was used in the rhythm distance measure to account for the relative weights or contributions of the prior $\hat{P}^{(i)}(l)$ of the timbre channels to the mixture of timbre channels. This form was used because lower-energy timbre channels will have less effect on similarity between two grooves than higher energy timbre channels. The impetus for this design choice comes from research (Kung et al. 2011) that examined the effects of stress and accent on listener judgments of rhythm similarity.

4.2.5 Results

Figure 4.4 shows the rhythm similarity measure for groove retrieval produced the best overall precision, with the average Euclidean distance performing slightly worse but still better than either sub-band method (Qingyuan et al. 2011). The sub-band method with rhythm weighting performed considerably worse than the average Euclidean distance, suggesting the sub-bands increased confusion in the feature distances in the weighted form. This observation follows the design intuition behind H-PLCA because H-PLCA provides an implicit similarity heuristic in the timbre channel selection process. In other words, the sub-band method generates at a similarity space, whereas the H-PLCA method produces a context-based similarity space with reduced background noise.

CHAPTER 5

LATENT STRUCTURE ANALYSIS FOR MUSIC COMPOSITION

Spectrally decomposed features can be used to produce novel music (Casal and Casey 2010, Osetinsky 2010). While the techniques for source separation have been available for more than 12 years, the non-scientific use of spectral decomposition algorithms has received little attention until recently (Casey and Westner 2000). This lack of interest may be the result of limited access to these algorithms by artists and the general lack of interest in the MIR and signal processing communities to explore non-scientific applications. Although it may be argued that spectral decomposition is not an intuitive process, it is suggested in this dissertation that the process is intuitive.

The characterization of spectral decomposition techniques as source separation leads to misconceptions about what the algorithms actually do and how the parameters affect the process. For this reason, section 5.1 explains how the parameters affect the compositional results in the featured algorithm, PLCA. In the following section, four works employing PLCA are described in order of their first performances. Each work manipulates components differently, but they all use components to articulate specific and recognizable characteristics of the original audio samples at important moments in the compositions.

5.1 A Spectral Decomposition Repertoire

The PLCA2d algorithm used in the current version of SoundSplitter can extract fixed, recurring patterns in frequency and amplitude. These components retain a qualitatively higher level of the original structure of the audio than nonpattern extraction methods (e.g., bandpass filtering, spectral frequency decomposition, or phase-vocoder decomposition). The following sections discuss five works employing these techniques in the order of their first performances.

5.1.1 *Strange Charmed*, by Michael Casey and Simon Atkinson

Strange Charmed *Strange-Charmed* (1999) is the earliest spectral decomposition-based composition, and the composers Casey and Atkinson used an independent subspace analysis (ISA) of spectrogram data (M. Casey 1999, Casey and Westner 2000). The source material for this work is derived from natural sound recordings captured on a visit to Cambodia. The decomposed sources are recombined in a way that draws on the legacy of musique concrete, which was pioneered by Schaeffer in Paris in 1948 (Kim-Cohen 2009). Aesthetically, the work draws on the dissociation from the original sources using grain-like presentations of the sources, but what emerges is, in effect, different timbral streams of continuity. As the work progresses, the sources become more distinctive, identifying themselves as exotic birds, insects, and flowing water. The process of revealing–a common trope in the works presented here—supplies the context for the disassociated sounds heard at the beginning of the piece.

5.1.2 *Stratovinksy*, by Paul Osentinksy

The concept behind *Stratovinksy* (Osentinsky 2010, Osetinsky 2010) the gradual revealing of an important musical instance: the iconic, jaggedly repeating chord from the "Omens of Spring: Dances of the Youths and Maidens" movement in

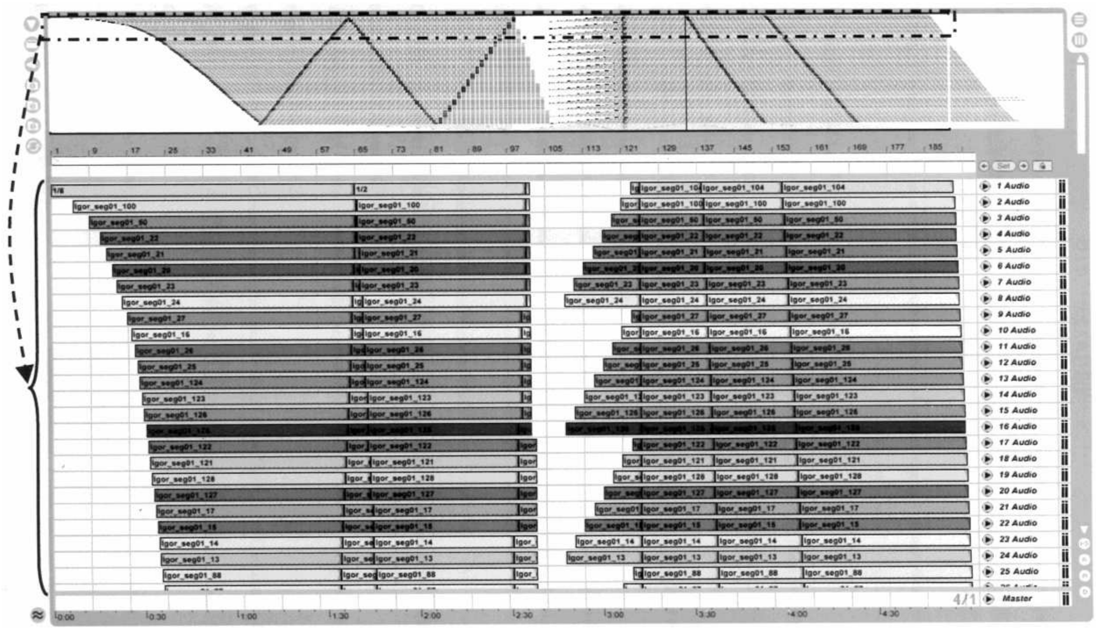


Figure 5.1: Osentinksy components introduced one at a time, with the misaligned components gradually aligned to bring the heterophony of asynchronous components into a state of order, clearly revealing the source as Stravinskys iconic Rite of Spring chord (Stravinsky 1989).

Igor Stravinskys famous *Rite of Spring* (1989). Using PLCA, 128 components were extracted, temporally misaligned, and distributed in 16 virtual channels across a quadraphonic array using sound spatialization. The layered components gradually came into alignment over the course of several minutes, illustrated by the relationship illustrated in Figure 5.1. The effect of this process is similar to seeing a blurred object slowly come into focus.

The component extraction was set at a relatively high parameter of 128, and the composer introduced sound spatialization techniques because competition between spatial and frequency queues has the tendency to create auditory illusions, a concept borrowed from Bregmans (1994) ASA. While the component separation in this piece is relatively simple and the *Rite of Spring* chord struggles throughout the work to realign, the result is particularly effective.

5.1.3 *Decomposing Autumn*, by David Plans Casal

Decomposing Autumn (Casal 2008a;b) examined live improvisation using audio matching of an input signal from a custom-designed guitar, with components extracted from Ligetis Autumn in Warsaw (2010). This work builds on Casals previous work that combined audio database matching with music improvisation (Casal, 2008). The piece is innovative because it combines latent component analysis with music information retrieval using audio matching. Casal later extended this concept to include crowd sourcing, in which people could visit a website to add their own vocal or instrumental rendition of different resynthesized components (Collins 2011).

5.1.4 *Violine*, by Spencer Topel

Violine (Topel 2010), a piece for solo violin and laptop, was, *Stratovinksy*, composed to highlight and reveal musical structure. The source material for each movement consisted of three short (i.e., 12 to 90 second) audio clips of J. S. Bachs "Chaconne in D Minor" from Partita No. 2 for solo violin (see Figure 5.2). Doing so preserved not only the composed structure explicit in the Bachs notation, but also the timing and articulation supplied by the performer (e.g., rolling of chords, chord voicing, and rubato).

The approach in this composition was two-fold: (a) use SoundSplitter to perform a decomposition to isolate individual notes or pitch classes and (b) use

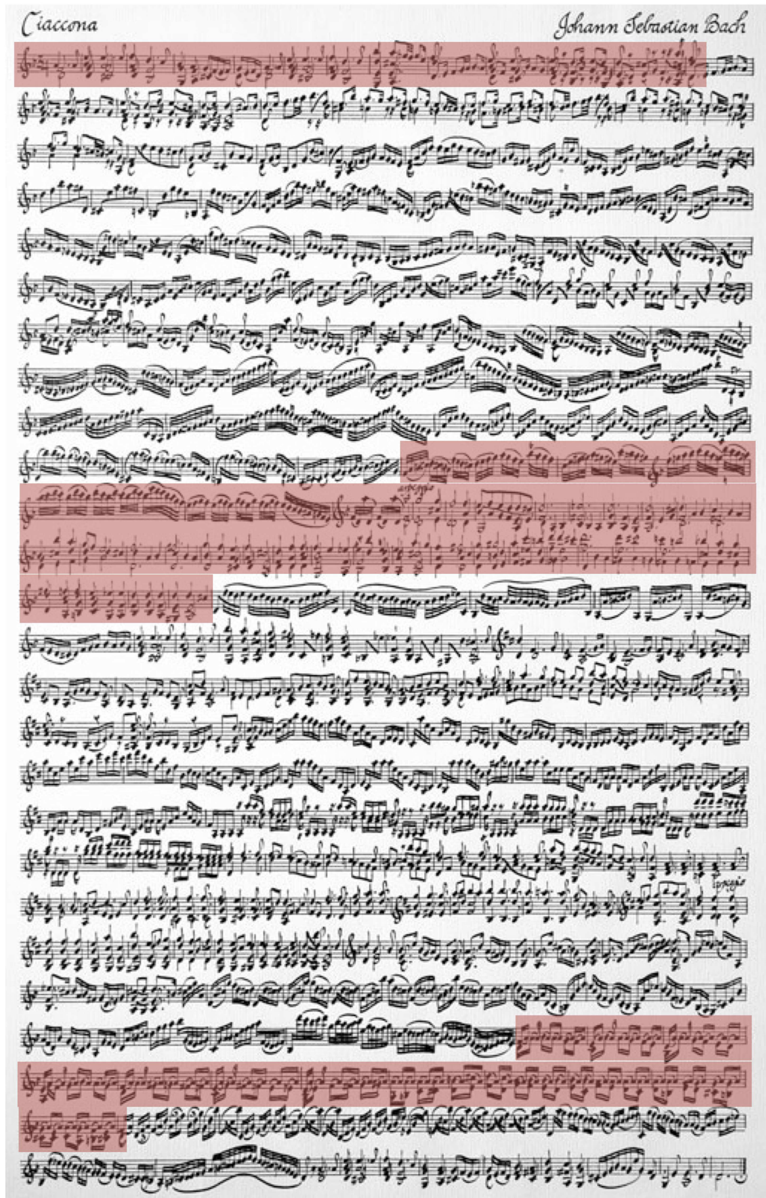


Figure 5.2: Three sections used as source material (highlighted in red) for *Violine*, from J.S. Bach's "Chaccone in D Minor" for solo violin

SoundSpotter to match a live violin signal directly to audio features analyzed on the extracted component database (Casey 2011), which is similar to the techniques used by Casal and Casey (2010). The basic function of SoundSpotter enables pitch and timbre characteristics to be matched, both of which were used in Violine. By using SoundSplitter and SoundSpotter, the compositional material becomes a combination of the composers intentions and the performers interpretation, similar to the pieces discussed in Casey (2009).

The component decomposition parameter was critical in the pre-composition phase of this piece. An eight-component decomposition proved to be effective for extracting clear, well-formed sounding components. A musical score was then written using the analysis provided by the SoundSplitter decomposition. A recent version of SoundSpotter as a VST plug-in provided an immediate and interactive way to match the sounds of the live violin, resulting in a near-seamless counterpoint between the extracted components and the composition.

5.1.5 *Intersecare*, by Spencer Topel

Intersecare (2011) extends the concept of *Violine* across the entire collection of Bach Cello Suites. Inspired by the drawing (after Sleeper 06 by Chloe Piene) shown in Figure 5.3, with its strand-like fragility intersecting in often-dramatic ways, this work attempts to build shape using similarly limited and sparse means (Topel 2011b).

Feature matching in the form of pitch-timbre matching using SoundSpotter allowed the performer to activate and then musically respond to emergent pat-

Figure 5.3: “after Sleeper 06” by Chloe Piene. Charcoal on vellum. 43.75 x 34 inches, courtesy the artist.

terns in resynthesized components. Furthermore, the unique timing of the performance provided rich variation in micro-timing information, creating a sense of undulating forms that echoes concepts described in (Beaudoin 2009).

Like *Violine*, an eight to ten component decomposition was sufficient for extraction. Figure 5.4 shows the first page of the score. Here the player has the choice of navigating different pathways through specific pitch material that overlaps with that of the components. There is inherent rhythm in each strain (Osetinsky 2010), and as a result, the performer is required to imitate the rhythm while articulating the pitch structure by moving between the different pathways indicated in the music notation in a structured-improvisation fashion.

Like *Violine*, eight to ten component decomposition was sufficient for extraction. Figure 5.4 shows the first page of the score. Here the player has the choice of navigating different pathways through specific pitch material that overlaps with that of the components. Since there is inherent rhythm within each "strain" (Osetinsky 2010), the performer is required to imitate the rhythm, while articulating the pitch structure by moving between the different pathways indicated in the music notation in a structured-improvisation fashion.

5.1.6 *Elementary Sources*, by Spencer Topel

Elementary Sources (Topel 2011a, Topel and Casey 2011b) examines the different roles of correlated timbral features in a single source. In other words, unlike the previous works described above, the source material is recognizable throughout the work. The independent component analysis was, in essence, a timbral separation because the PLCA algorithm was used not to decimate the sounds

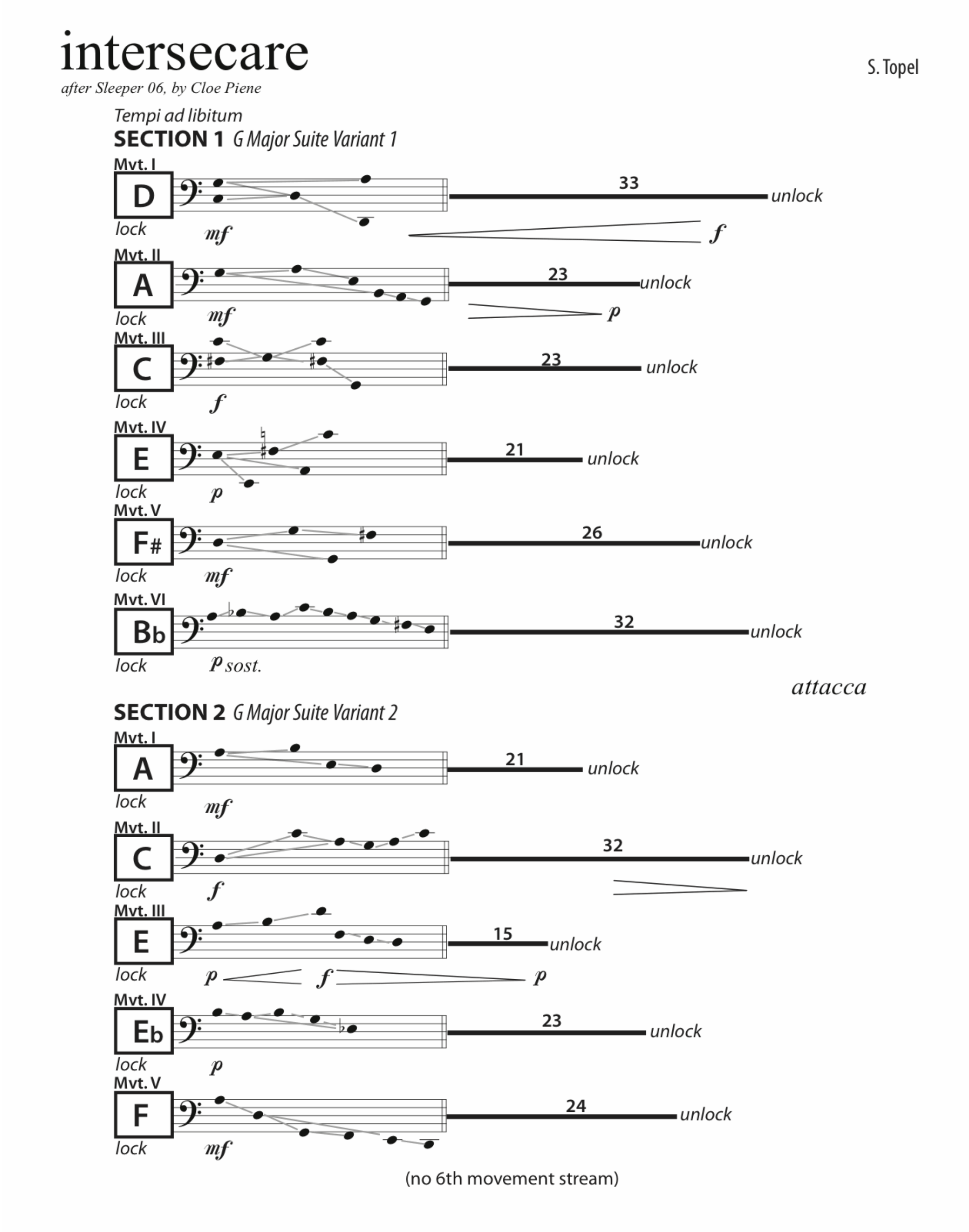


Figure 5.4: The first two sections of *intersecare*, where each line represents a different set of resynthesized components, and the ligatures between the indicated pitches are pathways generating motivic cells. The boxed letters are the activation pitches for a given line.

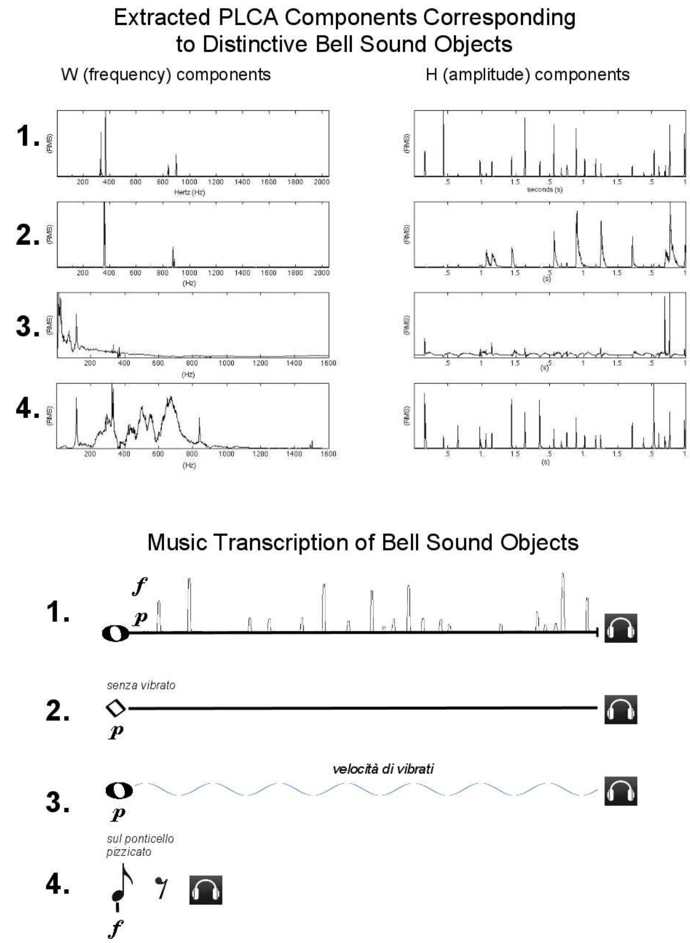


Figure 5.5: Each horizontal row corresponds to an extracted component, where the left-hand side of the figure shows the transcriptions to music notation, and the right-side shows the plots of the H and W decompositions per-component. From top to bottom: 1)A transient-laden component over the relative duration of a whole-note, where the peaks represent loud articulations and the troughs are equivalent to pianissimo. 2) A clean bell-partial, 3)“wobbling” of the bell-halve settling on the concrete. 4) Cement “click,” articulated by a near-pitchless pizzicato near the bridge of the instrument.

into unrecognizable objects, but to decompose the audio into different timbral objects that contribute to the identity of the original source.

The movement described in this dissertation was written using a single audio recording of brass bell-halves dropped on concrete. PLCA was used to extract components related to different events segregated by timbre, which included cement clicks caused when the metal hit the ground, inharmonic partials from the metal as the bell-halves rang, and the oscillation, or wobbling, of the halves as they came to rest on the cement. Different component extraction parameters were explored, and the components were resynthesized and auditioned. Through trial and error, an eight-component extraction was identified as the best sounding results. Additional experimentation with component re-synthesis included creating hybrid components that had cross-characteristics with the three categories described above. Specifically, the *h* and *w* components from PLCA were swapped with one another to extend the timbral palette without adding new material.

A central idea across the entire work was to acoustically synthesize a specific audio sample, like the bell-halves, with an entirely different set of sources, such as a string quartet. This is not unlike the concept Gérard Grisey (1975) presented in his landmark work *Partiels*, in which he describes the concept of using the orchestra for the purpose of macro-synthesis. In this macro-synthesis, each instrument of the orchestra contributes specific time and frequency information rather than distinctive sections and instrumental motives.

Two methods were explored to determine the best way for the four acoustic instruments to perform the different extracted components. First, a notation for each component was devised to best articulate the timbre content, described in section 5.5. Second, audio samples were provided for the performers to audition the sounds for themselves in order for them to determine the execution of these components.

The two methods proved to work better in combination than in isolation because the notation provided a starting point and the quality of the playback of the components was high enough to provide additional information for each performer. With the basic elements of the bell-halves translated to the quartet, it was now possible to use a combination of live electronics and sample playback to achieve a fairly close relationship between the sampled sources and the acoustic instruments. The result was the creation of an interstitial space between the bell samples and their transcriptions where relationships in the compositional materials were a subsequent outgrowth of the timbral features from the original bell-halve sources.

5.2 Summary

This chapter presents recent compositions using the PLCA2d algorithm encapsulated in a tool for composition and improvisation called SoundSplitter (Topel and Casey 2011b). The control of timbral reference is a striking similarity among these works, specifically, the control used to reveal or obscure the sources in terms of their morphological archetypes (Thoresen and Hedman 2007). However, each composer discussed above demonstrated a different style of spectral decomposition, suggesting the combination of such morphologies is highly particular to the source audio and the structures in the audio that interest the composer.

Another way these works differ from the spectromorphology paradigm is in the preservation of the event structure. In particular, there is increased continuity with respect to the inherent component temporal structure, which is true on the micro-level in *Stratovinsky* and *Decomposing Autumn* and at a higher level in *Strange Charmed* and the works presented by the author. Inherent temporal structure, as it relates to micro-timing (Beaudoin 2009), also extends the concept of structure in spectralism because it captures correlated spectro-temporal events across arbitrarily large time scales.

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

6.1 Contributions

This dissertation examines the application of spectral decomposition in music analysis and composition, and it is noted that the latent structure of music can be accessed using a technique called PLCA. The idea of latent structure is conceptualized in creative and research applications. In creative applications, latent structure analysis is an extension of spectralist techniques because spectral components and spectral decomposition components access low-level auditory features and reveal correlations between reoccurring frequency and time components. In research applications, latent structure is contextualized through timbre channels and shows that relevant content is found in these channels and information outside of these channels is noise. For the purpose of the groove retrieval used in the present study, a hierarchical version of PLCA called H-PLCA revealed timbre channels containing background rhythmic information consistent across all the tracks in a database. The contributions of this study include the following:

- **Latent structure sources:** This dissertation claims there is a link between spectralism and spectral decomposition. A technical and aesthetic lineage with spectralism is intuitive because both spectral decomposition and spectralism use sub-perceptual information in sound sources. Yet, it is the access to reoccurring patterns that extends spectralist techniques. This not only allows for new ways to produce music, but also to analyze it. As stated in chapter 1, access to repetition of spectral patterns means that time scale and frequency scale manipulations are simply one of many options because the temporal characteristics of source materials can be kept in tact. The contrast between compositional approaches described in chapter 5 illustrates this observation.

- **Spectral decomposition contextualized by principles of unconscious listening:** Using Bregmans (1994) ASA as a starting point, underlying traits of primitive segregation are discussed with respect to spectral decomposition. It is proposed that PLCA reveals more information than CASA methods. Finally, it is argued that spectral decomposition reveals a precognitive layer of music information that is hidden from listeners.

- **Latent structure analysis:** Using PLCA as a reference, this study shows that an earlier form of latent structure analysis has been used by spectralist composers, such as Dufourt (1979), Schaeffer (1967), and Risset (Risset and Wessel 1982), to access sub-perceptual musical information. A general framework for digital musicology-based analysis is proposed to address challenges posed by analyzing such music with examples of two studies using latent structure analysis: (a) *Partiels Pour 16 Ou 18 Musiciens* by Grisey and (b) the fourth interlude from Cages *Sonatas and Interludes for Prepared Piano.*

- **ISHKUR, a newly annotated dataset available to researchers:** Rhythm similarity contextualized by timbre defines a specific rhythmic background called a groove. A groove retrieval analysis with ground-truth annotations (Qingyuan et al. 2011) was conducted on a dataset called ISHKUR and reveals more information about timbre channels than subband analysis.

- **Spectral decomposition analysis for composition, a repertoire:** This study has identified a small corpus of musical compositions that used PLCA to access structure and compositional material. These compositions span 10 years, and they illustrate innovations in component separation tailored to specific goals in the compositions. In contrast to source separation research, this study shows that rank estimation is an important free parameter in the creative process, and the use of this parameter depends on the context of the component extraction.

6.2 Future Work

The results of this study also have implications for future work:

- **Utilizing shift-invariant spectral decomposition for creating music:** The decomposition methods discussed in Chapter 5 use the non-shift-invariant of PLCA. That is, the extracted components resemble fixed frequency information useful for accessing reoccurring spectral patterns that do not move in the spectral domain, such as individual pitches or fixed percussion timbres. Shift invariance allows for the extraction of patterns that occur in pitched instruments and non-fixed percussion (e.g., marimba, xylophone etc.)
- **Improvement of component extraction methods:** One of the most important attributes of spectral decomposition is its ability to extract independent features within the spectrum. All of the algorithmic tools discussed in this dissertation operate on the assumption that the spectral features belong to one component or another but are not shared between two com-

ponents. Alternatives to PLCA use state-of-the-art statistical methods to perform source separation and allow exchangeability of features.

Several directions for the application of shift-invariant PLCA in composition and improved extraction techniques are identified below. A few of these methods are already being used in composition and research, while other methods are speculative.

6.2.1 Scale and Shift-Invariant Spectral Decomposition

Non-shift-invariant PLCA is effective for decomposing consistently recurring, fixed spectral shapes, such as single-pitch classes or percussive instruments with fixed timbres. With sources with greater variation in timbral behaviors, such as an orchestra or human voice, shift-invariant PLCA (SI-PLCA) (Shashanka et al. 2007) or scale-invariant PLCA may be preferred (Hennequin et al. 2011) because the SI-PLCA family of algorithms is designed to handle the extraction of patterns that shift in the spectrum, for example, instruments such as the piano may have recognizable and consistent spectral and amplitude profiles that shift in the spectrum as a result of a change in pitch. To illustrate this, Figure 6.1 shows a spectrogram of the first 10 seconds of the song *Because* by the Beatles.

Figure 6.2 illustrates a SI-PLCA decomposition. Unlike a PLCA component extraction, the SI-PLCA components reflect the preservation of the instrumental scale consistent with the timbre of that instrument. In a PLCA extraction, each note or pitch-class would be extracted as an independent component(see section 3.2.1).

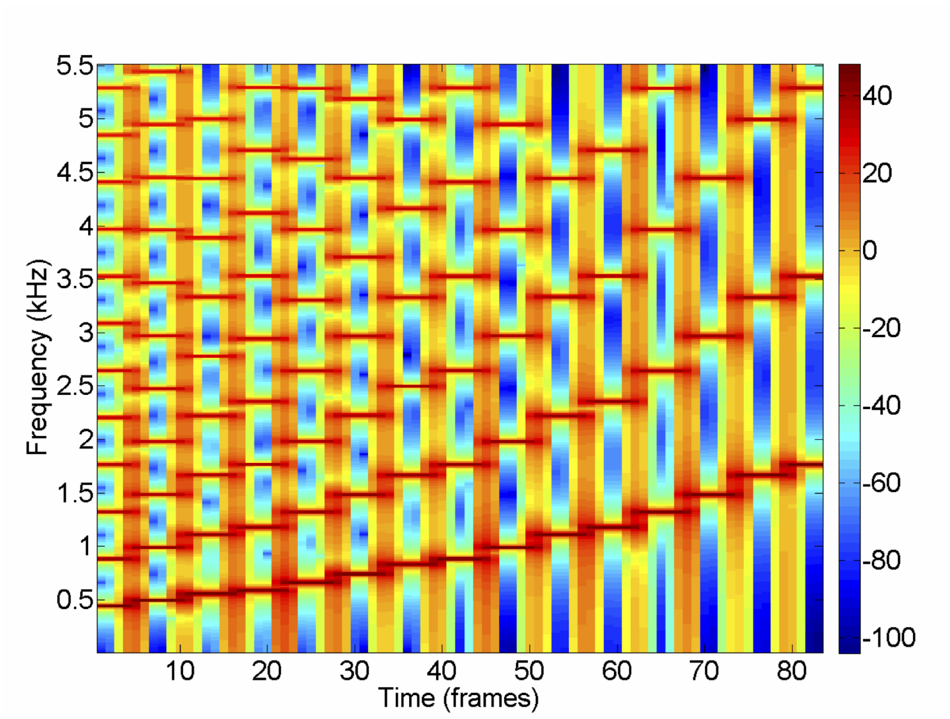


Figure 6.1: The original spectrogram information for the song *Because*, by the Beatles (Hennequin et al. 2011).

6.2.2 Future Directions in Spectral Decomposition Methods

A departure from techniques related to probabilistic spectral decomposition, such as PLCA, are Non-Parametric spectral decomposition techniques, which offer ways of extending spectral decomposition methods, either by introducing ways of estimating components or by improving the quality of extracted components.

Instead of starting with a model that assumes a fixed number of components, (i.e. the number of marginals to extract, which in PLCA2d is a fixed parameter), such models either start with a large number of components, and remove redundant or non-contributing components below a fixed threshold (Hoffman

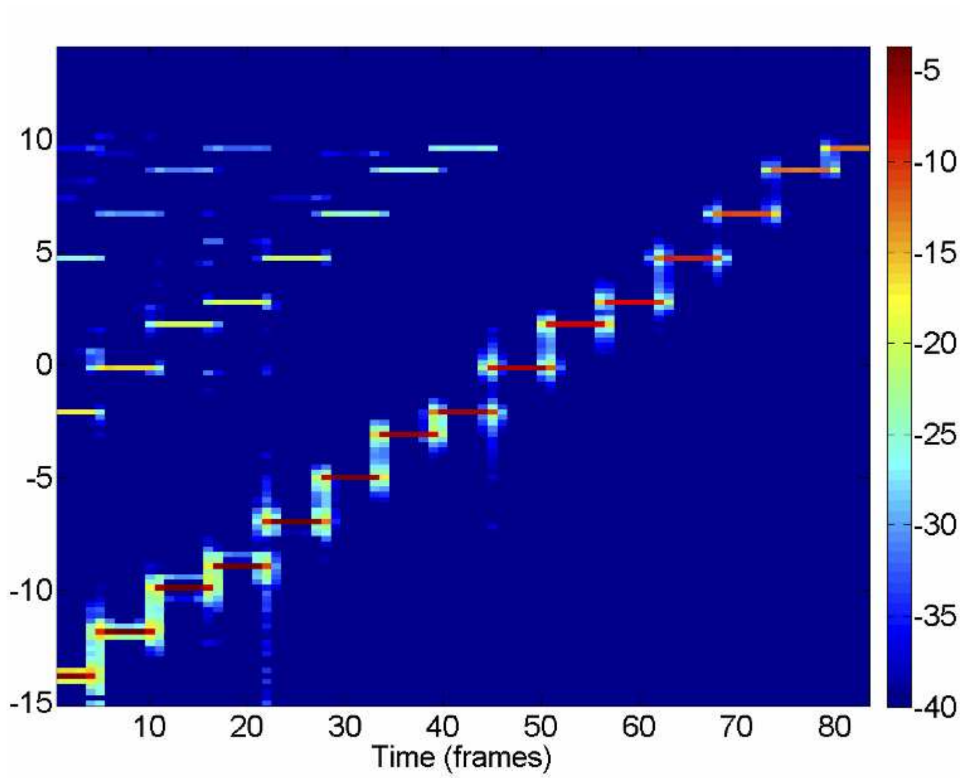


Figure 6.2: An impulse distribution of semitone information extracted using SI-PLCA from the Beatles' song *Because* (Hennequin et al. 2011).

et al. 2010) (Weiss and Bello 2010); or they start with a finite number of components and provide constraint for the limit as the number of features approach infinity (Griffiths and Ghahramani 2006).

6.3 Non-Linear Latent Structure Analysis

Another analysis method that aims to identify structure in music using spectral features belongs to a family of techniques called concatenative synthesis. This type of music processing is related to pitchsynchronous overlapped (PSOLA) speech synthesis. This system concatenates prerecorded speech phonemes with perceptually based units constrained by spectral-match quality using the audio features and temporal well-formedness of selected units. In essence, PSOLA attempts to synthesize human-sounding speech. The musical equivalent of PSOLA is Caterpillar (Schwartz 2004), one of the earliest systems to employ concatenative synthesis techniques. The source database consisted of short audio samples organized by pitch, range, and timbre. Using the well-formedness constraints of concatenative synthesis, Caterpillar assembles sequences of samples based on the best matches to either a symbolic (notes) or synthesized audio input file.

The first concatenative-based model to perform matching on sequences of audio features was SoundSpotter, developed at the Mitsubishi Electronic Research Laboratories in 2001 (Casey 2009). This first-generation SoundSpotter was intended as an audio segment retrieval-by-similarity system and built on the MPEG7 audio descriptor set and, like Caterpillar and Musaicing, was non-real time. In 2003, a live audio input replaced the query selection tool in the first SoundSpotter, and the new version of the system was able to run in real time. Furthermore, instead of using a Didden Markov model to generate discrete states, the newer system performed matching directly on audio features (Casey 2011; 2009).

With respect to latent analysis, SoundSpotter enables components to be matche to an arbitrary input signal using timbre features. The implications of this idea reach beyond the mosaic-like techniques discussed above and point to ways that timbre can manipulated with a wide range of expressive possibilities, as demonstrated in the compositions using SoundSpotter discussed in chapter 5. SoundSpotter can also be used to overcome problems, such as devising a shift-variant PLCA, by simply collecting all linear PLCA2d components as a single index of continuous audio and using the non-linearity of SoundSpotter to access these components through some query or input signal.

Non-Parametric Methods: Gamma Process Nonnegative Matrix Factorization

The goal of a non-parametric spectral decomposition model is to circumvent the problem of component estimation. Hoffman (2010) described a model called gamma process nonnegative matrix factorization (GAP-NMF) that extends the design of the signal model described by Abdallah and Plumbley (2004). Similar to PLCA and previous spectral decomposition algorithms (i.e., the spectrogram $X = WZH$), GAP-NMF replaces the diagonal Z with θ, which describes a hidden vector of non-negative values where θ_l corresponds to the total amplitude or gain of each component l. The GAP-NMF model then allows for a large number of components L to be reduced down via a sparse prior on θ, which biases the model keep only contributing components. This process is essentially the same rationale behind IS-PLCA and H-PLCA. The main difference being that GAP-NMF utilizes a generative process, paraphrased from Hoffman (2010):

$$
\begin{aligned}
&W_{ml} \sim Gamma(a, a) \\
&H_{ln} \sim Gamma(b, b) \\
&\theta_l \sim Gamma(\alpha/L, \alpha c) \\
&X_{mn} \sim Exp(\sum_l \theta_l W_{ml} H_{ln}).
\end{aligned} \tag{6.1}
$$

where θ approximates an infinite sequence drawn from a gamma process with an α shape parameter and inverse-scale parameter αc (Kingman 1993). As Hoffman (2010) correctly pointed out, the Gamma process and Dirichlet process are closely related, and when a gamma process is normalized by weighting its atoms to 1, it is a Dirichlet process. A significant difference between this non-parametric approach and IS-PLCA and H-PLCA is that the computational problem centers on the posterior inference and uses variational inference instead of the EM algorithm (Hoffman 2010).

Priors on Amplitude-Time Components

In the hierarchical PLCA used for groove retrieval, timbre channel extraction required adapting discovered W priors to local features (Qingyuan et al. 2011). Alternatively, there may be ways of using H priors to perform rhythm and structure analysis, which is essentially how IS-PLCA uses h information to iteratively determine segmentation boundaries (Weiss and Bello 2010).

Another approach using temporal priors would be envelope prediction and matching, which means that instead of having global timbre channels across a dataset there would be global envelope channels describing amplitude characteristics. This could allow for more sophisticated temporal modeling using a priori information for segmentation and frame alignment (Rhodes et al. 2006).

Clusters of Classes of Latent Features

A current statistical method called the Indian buffet process (IBP) could also be used to make features exchangeable and inference tractable (Griffiths and

Ghahramani 2006). The authors liken this behavior to analyzing sequential images, such as video, in which the more images that are presented, the more objects are discernible in a given frame. A similar analogy can be made for analyzing music audio, either by a machine or a human listener. In this case, objects are correlated according to time-frequency features that are either reinforced by similar patterns that arise over some arbitrary amount of time or undermined by the non-reinforcement of these patterns. The IBP model proposed by Griffiths and Ghahramani (2006) starts with a few features and expands that number toward infinity, which is different from other non-parametric Bayesian models for music audio decomposition, notably GAP-NMF (Hoffman et al. 2010). GAP-NMF starts with many components and reduces them to a few components. IBP is attractive for music analysis because it retains maximal variation in the combinations of features while permitting features to be shared across different components.

Doshi-Velez and Ghahramani (2009) provided a better definition of sharing components and proposed a hierarchical framework for IBP by evoking the Dirichlet process (DP) and the Chinese restaurant process (CRP). The authors pointed out that because DP and CRP are distributions on discrete and clusterable distributions (Teh et al. 2007) CRP can be represented in matrix form by setting $c_{nl} = 1$, if observation n belongs to cluster l. The IBP is a model in which each observation is associated with $Poisson(\alpha)$ features. In the general description of their model, DP-IBP draws C from a Chinese Restaurant Process (CRP) (Blei et al. 2004), and M from IBP, where $f(c_n^\top m_k) = c_n^\top m_k$, and thus $Z = CM$:

$$
\begin{aligned}
&C \sim CRP(\alpha_C) \\
&M \sim IBP(\alpha_M) \\
&z_{nk} = c_n^\top m_k
\end{aligned}
\tag{6.2}
$$

where given N total observations, the probability that a signal observation, n, contains an active feature, k, is $\frac{r_k}{N}$, and where r_k is the number of observations currently using feature k. Building on the CRP analogy, the authors suggested that instead of single dishes the dishes are drawn from meals, and different meal combos share specific dishes (features). The dishes themselves are drawn from some discrete base distribution, a method also used in PLCA to initialize the algorithm. They argued that the combined properties of DP and IBP ensure DP-IBP is exchangeable over features and observations, something that is impossible to do with PLCA2d.

The experimental results for DP-IBR include a series of toy problems examining the effectiveness of recovering features and clusters of these features. One test of possible relevance was the synthetic block test. In this task, Doshi-Velez and Ghahramani (2009) generated 700 images with 6x6 pixels. These images contained four types of blocks, shown in the lower-left corner of Figure 6.3. The experimenters then performed the synthetic block test using Gibbs sampling (Carter and Kohn 1994) with the parameters described by Doshi-Velez and Ghahramani (2009). The results show an accurate recovery of the underlying features and cluster of features, with the Gibbs sampler quickly converging near the true number of features and clusters.

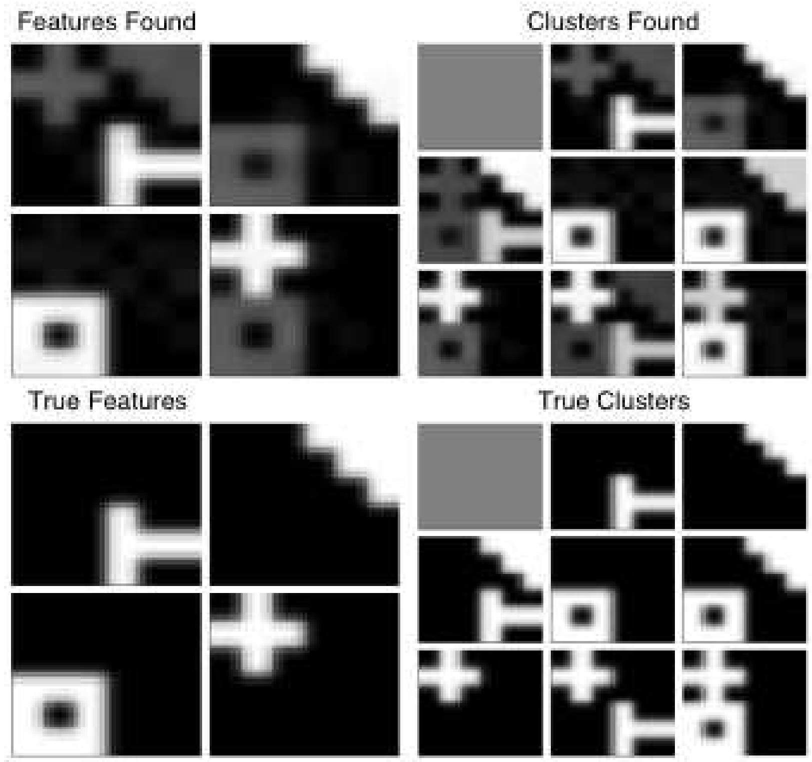


Figure 6.3: A result from a synthetic block experiment testing, where four types of features and nine types of clusters were accurately recovered using DP-IBP. The features and clusters recovered from DP-IBP (top row), clearly match the underlying structure (bottom row) (Doshi-Velez and Ghahramani 2009).

In audio, signal features cannot be hierarchically clustered, and this is the null hypothesis for this model. However, there is evidence in the speech literature that it is possible to discover structure using hierarchical clustering methods similar to the method illustrated in 6.3, in particular, the Sticky HDP-HMM model used to determine from the audio who is speaking when there are multiple speakers having a conversation, that is, the speaker diarization problem (Fox et al. 2009).

6.4 Conclusions

This dissertation examines music audio analysis, composition, and retrieval using spectral decomposition techniques. The results of a number of different music analyses show that spectral decomposition methods may increase access to latent structure (outlined in chapter 1). While there is still work needed in the area of spectral decomposition algorithm design, current algorithms and computing power can be meaningfully applied to a broader set of music-related analysis tasks.

This investigation has also highlighted that the spectral decomposition of music audio can be used for more than source separation. In addition, the results suggest there are connections to past ideas on the organization of sound and spectral decomposition, previously discussed as timbre, specifically, in the work of Schaeffer, spectralist composers Gérard Grisey, Tristan Murail, and Hugues Dufourt, and the research of Jean-Claude Risset, the Bregman Music Audio Research Studio, and the MIR and signal processing research communities.

It is possible to envision spectral decomposition methods evolving out of state-of-the-art statistical methods (Doshi-Velez and Ghahramani 2009), which would provide near-limitless access to correlated patterns in audio with production-level quality. In lieu of these inevitable discoveries, the range of applications with current techniques illustrated in this dissertation shows that spectral decomposition is relevant to a much wider range of applications and provides us with a new way to access music through the analysis of latent structure.

APPENDIX A

GLOSSARY

- **Auditory Scene Analysis (ASA)**:The model of the human biological auditory system proposed by Albert Bregman (Bregman 1994), which explains how auditory processing occurs at different psychoacoustic stages.
- **Computational Auditory Scene Analysis (CASA)**:A family of computational models designed to mimic ASA in Machine Learning applications (Shao and Wang 2008).
- **Chinese Restuarant Process (CRP)**: A finite mixture model incorporating exchangeability, where latent variables are distributed according to partitions defined by a random process (Blei et al. 2004).
- **Dirichlet Process(DP)**: Is a stochastic process used in models of data utilizing Bayesian nonparametric frameworks. It is sometimes referred to as a "distribution over distributions", which means that each draw from a Dirichlet process is itself a distribution (Teh 2007), and has been compared to the Chinese Restaurant Process (Teh et al. 2006).
- **Fast Fourier Transform (FFT)**: Used to acquire spectral content, with no account for the temporal activation of events. It is represents frequency content only, over a finite or infinite window of time. This transform is essentially a histogram, with parameterized bin sizes designating the resolution of the spectrum.
- **Mel-Frequency Cepstral Coefficients (MFCC)**: A robust timbre feature that also captures event information, (e.g. rhythm), but is insensitive to pitch. This method is often used to for copyright purpose because of its ability to do highly-accurate recording matching (Mermelstein 1976).

- **Hierarchical Probabilistic Latent Component Analysis (H-PLCA)**: An iterative algorithm designed to extract a background model consisting of a finite number of timbre channels. The hierarchical term refers to the process by which global basis functions are fitted to local features (Topel and Casey 2011b).
- **Indian Buffet Process (IBP)**: A modification to the Chinese Restaurant Process for the purpose of infinite mixture models. Indian Buffet Process is itself a stochastic process that allows partitions of latent variables to include partitions of "draws" of variables defined by a prior on infinite matrices (Griffiths and Ghahramani 2006).
- **Iterative Segmentation-PLCA (IS-PLCA)**: Another variant of PLCA, IS-PLCA is designed with the intention that music has naturally-occurring points of segmentation, (e.g. phrases, resets, etc.) (Weiss and Bello 2010).
- **Probabilistic Latent Component Analysis (PLCA)**: Components can be "extracted" from an audio mixture resulting in a representation that shares correlated features between the frequency-time and amplitude-time domains allowing for well-formed structure elements to be isolated, recombined, etc. It is designed to operate on a magnitude-only time-frequency representation, such as the magnitude- STFT (Shashanka et al. 2007).
- **Probabilistic Latent Component Analysis Implimentation (PLCA2d)**:The name of the implemented PLCA algorithm, PLCA2d, uses a 2 dimensional kernel density estimation (Venables and Ripley 2002), to compute the correlations between frequency and amplitude marginal probabilities for a given component, described in (Shashanka et al. 2007).
- **Non-Negative Matrix Factorization (NMF)**: Is a genre of techniques use in spectral decomposition of positive-only (magnitude) matrices. When the

found non-negative matrices are multiplied together, they approximate the original non-negative matrix (Lee et al. 1999).)

- **Shift-Invariant Probabilistic Latent Source Separation (PLCA)**: An extension to PLCA that allows for convolutive bases (Smaragdis et al. 2008), allowing for normalize *w* basis functions.
- **Short-Time Fourier Transform (STFT)**: Capable of capturing both spectral content and temporal events. This is made possible by performing iterative sampling over time with a fixed window, and what is called a ”hop”, where the window is moved to an overlapping position on a previously sampled window. Using a smoothing kernel, many overlapping windows result in a STFT spectrogram with changing amplitudes for the analyzed frequencies.
- **singular value decomposition (SVD)**: is a factorization of a real or complex matrix P that can be decomposed as the product $P = UDV^T$ where the singular vectors in V and U are orthonormal. This is a widely used method in statics and machine learning; first proposed in (Golub and Reinsch 1970).

APPENDIX B

EXPERIMENTAL CODE EXAMPLES

B.0.1 Analysis of *Interlude No. 4* by John Cage

```
%________________________________________________________
%
%Analysis of the Interlude No. 4 for Prepared Piano by John Cage.
%________________________________________________________
%
%by Spencer Topel
%%
%Generate PLCA Components from first twenty seconds of "Fourth ...
    Variation"
%iteratively across 8, 12, 20, and 32 components and Resynthesize:

%number of components to extract
numComponents = 20;

s = wavread( 'fourthVariation.wav')';
filestem0 = 'section_1_';



% Go to time-freq via STFT with a window of 2048 and a hop of 1024
f = stft( s, 2048, 1024, 0, 'hann');


% Do PLCA2d Analysis
[w,h,z] = plca2d( abs( f), numComponents, 200, 0, 0, 0, [], [], ...
```

```
    [], 1);


% Resynthesize components
fn = abs( f) ./ (w*diag(z)*h);
fp = f ./ abs( f);
fp = fp ./ repmat( linspace( .5, 10, size( f, 1))', 1, size( f, 2));

    for i = 1:size( w, 2)
        tf = (w(:,i)*z(i)*h(i,:)) .* fn;
        y(i,:) = stft( tf.*fp, 2048, 1024, 0, 'hann');
        wavwrite(y(i,:), 44100, [filestem0 ...
            num2str(numComponents(n), '%02d')
%'_' num2str(i, '%02d') '.wav']);
    clear y
    end
```

B.0.2 Analysis of *Partiels* by Gerard Grisey

```
% What can Spectral Decomposition tell us about a spectralist ...
    composition?
%_______________________________________________________
%
%Analysis of G. Grisey's "Partiels" Pour 16 Ou 18 Musciens
%_______________________________________________________
%
%by Spencer Topel
%%
%Generate PLCA Components from first section of "Partiels" ...
    iteratively across
```

```
%8, 12, 20, and 32 components and Resynthesize:

%number of components to extract
numComponents = [8,12,20,32];

s = wavread( 'Section_1.wav')';
filestem0 = 'section_1_';


for n = 1:4
% Go to time-freq via STFT with a window of 2048 and a hop of 1024
f = stft( s, 2048, 1024, 0, 'hann');


% Do PLCA2d Analysis
[w,h,z] = plca2d( abs( f), numComponents(n), 200, 0, 0, 0, [], ...
    [], [], 1);


% Resynthesize components
fn = abs( f) ./ (w*diag(z)*h);
fp = f ./ abs( f);
fp = fp ./ repmat( linspace( .5, 10, size( f, 1))', 1, size( f, 2));

    for i = 1:size( w, 2)
latex
        y(i,:) = stft( tf.*fp, 2048, 1024, 0, 'hann');
        wavwrite(y(i,:), 44100, [filestem0 ...
            num2str(numComponents(n), '%02d')
%'_' num2str(i, '%02d') '.wav']);
    clear y
    end
```

```
end


  for i = 9:size( w, 2)
        tf = (w(:,i)*z(i)*h(i,:)) .* fn;
        y(i,:) = stft( tf.*fp, 2048, 1024, 0, 'hann');
        wavwrite(y(i,:), 44100, [filestem0 ...
           num2str(numComponents(n), '%02d')
%'_' num2str(i, '%02d') '.wav']);
    clear y
    end

%do autocorrelation on extracted components, trombone E2 ...
   (time-stretched), and

%original signal.


%use Euclidean distance to measure from extracted components and ...
   trombone E2
%note, and original signal.
```

APPENDIX C

EXPERT LISTENER MARKUP RESULTS

C.0.3 Instructions to Reviewers

1. Your task as an expert reviewer is to classify the grooves and beats of all 1138 song excerpts in a research database. To do so, play each song and determine the best category.
2. Categories can be selected by selecting the boxes below the title ″Beat Categories″ on the “BeatIDExperiment” patch. All songs should have a category. Please limit the number of categories as best as you can. Additional category boxes are available if this proves impossible.
3. Songs that don't have beats should go in the ″non-beat″ category, regardless of style, genre, etc.
4. Once you have finished the evaluation, or want to save your progress partway, you can save your current session using either the ″Save Auto″ or ″Save New″ buttons. ″Save Auto″ generates the file ″Results.txt″ on your desktop. ″Save New″ allows you to name the evaluation file yourself.
5. If you want to resume a previously saved session. Use the ″Load″ button to select the file from the location you saved it.
6. The ″Restart″ button loads a black evaluation file in the event you want to start over.
7. When you are finished, please send along the saved evaluation file along with descriptions of the categories that you decided upon in the process. This should be simple and to the point. Even one word category should suffice.

C.1 Reviewer Categories

Two expert listeners provided markups for the ISHKUR dataset corresponding to what the instructions stated above. The average time for the reviewers to complete the task was four days. Among the three reviewers, the average music education and/or experience was approximately 20 years.

C.1.1 Categories Determined by Listener *AS*

Reviewer AS is a Hip-Hop producer and independent author named Amir Sa'id (Sa'id 2011), and these categories correspond to the markup file *ASLIST.ascii*.

Summary

1 Boom Bap/East Coast (New York Rap Sound).

2 West Coast/G-funk.

3 Break beat.

4 Electronic Drum Machine.

5 Southern Bounce (Southern Rap Sound).

6 Electro Funk.

7 Miami bass.

8 Drum n' bass/grime.

9 Techno.

10 House.

11 R&B Hip Hop.

16 non-beat.

Detailed Description

Categories 6, 8, 9, 10, 11, *and* 16 *did not have additional details.*

1. **Boom Bap/East Coast (New York Rap Sound):** Beat style and sound that emphasizes knocking customized drums, filtered bass line samples, and sample chops. Beat style and sound established modern form of sampling.

2. **West Coast/G-funk:** Beat style and sound (based on the late 1970s P-funk sound of George Clinton and Parliament) that emphasizes use of live instrumentation, particularly distinct melody lines.

3. **Break beat:** Beat style and sound primarily based on/driven by sampled funk breaks. Typically, most (if not all) of the elements of break-beat beats are made up of whole sampled breaks of pre-recorded music, usually early funk breaks.

4. **Electronic Drum Machine:** Beat style and sound that prioritizes and is driven by samples of soul music (typically between 1965-1975). Soul-sample style and sound prioritizes capturing the meaning and emotional feel of the soul music it samples and references. By there nature, soul-sample

   beats, along with East Coast beats, tend to be the most conducive to complex (advanced) rap lyricism.

5. **Southern Bounce (Southern Rap Sound):** Beat style and sound that

   prioritizes the use of synth sounds and deemphasizes traditional sampling style. Southern bounce/trap prominently features synth brass, highly syncopated electronic (stock/preset) drum sounds, and the ubiquitous 808 (Roland TR 808)

   bass drum. Southern bounce/trap beats range in arrangement scope and complexity, which in ch. 6 of my book, The BeatTips Manual, I divide into three separate categories: "non-orchestral," "semi-orchestral," and "orchestral" (or epic).

6. **Electro Funk:** *no description provided*

7. **Miami bass:** An off-shoot of Electro Funk that became part of basis for Southern Bounce.

C.1.2 Categories Determined by Listener *ST*

Reviewer ST is the author of this dissertation, and these categories correspond to the markup file *STLIST.ascii*.

1 Break + Pulse (usually lowest sub-division level).

2 Swing Break/ New Soul Break / Rap and Hiphop.

3 Straight Break / No Swing.

4 Techno + Accented Metric Structures.

5 Erratic, Heavily Distorted, and Esoteric Rhythm Patterns (relative to general rhythm patterns).

6 Triplet/PolyRhythmic.

7 Gabber-Repeating Bass.

8 Pulse Trance and non-break Trance beats.

9 2-Step (beats on 1 and 3 respectively).

10 Short-short Long (Beastie Boys).

11 Funk/Disco/house beats (2-step dominating over break patterns).

12 Drum Chatter and Drum Machine (beat origami).

14 Amerifro Drumming (faux Voodoo and Bhata).

16 non-beat.

C.1.3 Categories Determined by Listener *JG*

Reviewer JG is a Swiss DJ and producer, who attended Dartmouth College as a Master's of Engineering Student in 2010, and these categories correspond to the markup file *JG_BeatCats.ascii*.

1 Hip-Hop (slow tempo).

2 drum n bass (amen break).

3 time signature 3/4.

4 time signature 4/4 house & various tempo.

5 other noise (faster stuff).

16 non-beat.

Bibliography

S.A. Abdallah and M.D. Plumbley. Polyphonic music transcription by non-negative sparse coding of power spectra. In *Proc. 5th Intl Conf. on Music Information Retrieval (ISMIR)*, pages 10–14, 2004.

K.R. Agres and C.L. Krumhansl. Musical change deafness: The inability to detect change in a non-speech auditory domain. In *Proceedings of the 30th annual conference of the cognitive science society*, pages 975–980, 2008.

A.V. Aho and J.E. Hopcroft. *Design & Analysis of Computer Algorithms*. Pearson Education India, 1974.

A. Arbo. Some remarks on /hearing-as/ and its role in the aesthetics of music. *Topoi*, 28(2):97–107, 2009.

C.A. Arrell. *Part I. Blackbirds and Argot: Part II. Pushing the Envelope: Art and Science in the Music of Gérard Grisey*. PhD thesis, Cornell University, Aug., 2002.

G. Assayag, C. Rueda, M. Laurson, C. Agon, and O. Delerue. Computer-assisted composition at ircam: from patchwork to openmusic. *Computer Music Journal*, 23(3):59–72, 1999.

T.L. Bailey, C. Elkan, San Diego. Dept. of Computer Science University of California, and Engineering. *Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers*. Citeseer, 1994.

Big Bang. Cold nights in cairo. Swanyard Records, 1989.

D. Barry, D. Fitzgerald, E. Coyle, and R. Lawlor. Drum source separation using percussive feature detection and spectral modulation, 2005.

C. Batteux. *Principles of translation*. Printed by Sands, Donaldson, Murray, and Cochran for A. Donaldson, 1760.

R. Beaudoin. Some account of the practice of microtiming and musical photorealism. *Unpublished manuscript*, 2009.

M.A. Bee and C. Micheyl. The cocktail party problem: What is it? how can it be solved? and why should animal behaviorists study it? *Journal of Comparative Psychology*, 122(3):235, 2008.

C. Bey and S. McAdams. Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2):267, 2003.

D. Blei, T.L. Griffiths, M.I. Jordan, and J.B. Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 16:106, 2004.

A.S. Bregman. *Auditory scene analysis: The perceptual organization of sound*. The MIT Press, 1994.

A.S. Bregman. Auditory scene analysis and the role of phenomenology in experimental psychology. *Canadian Psychology/Psychologie canadienne*, 46(1):32, 2005.

A.S. Bregman, J. Campbell, et al. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of experimental psychology*, 89(2):244–249, 1971.

E.O. Brigham and RE Morrow. The fast fourier transform. *Spectrum, IEEE*, 4(12): 63–70, 1967.

G.J. Brown and M. Cooke. Computational auditory scene analysis. *Computer speech and language*, 8(4):297–336, 1994.

James Brown. Funky drummer, 1969.

D. Bundler. Interview with gerard grisey. *20th Century Music*, 1996.

M.J. Butler. *Unlocking the groove: Rhythm, meter, and musical design in electronic dance music*. Indiana Univ Pr, 2006.

J. Cage. Prepared piano music volume 1, 2. *New York, London, Frankfurt, Leipzig: CF Peters Corporation*, 1940-47.

J. Cage. Sonatas and interludes. *played by Maro Ajemian*, page CD 700., 1995.

J. Cage, B. Berman, and Inc OverDrive. *Music for prepared piano*. Naxos American Classics, 2005.

Jean-François Cardoso. Source separation using higher order moments. In *Proceedings ICASSP-89 (IEEE International Conference on Acoustics, Speech and Signal Processing)*, pages 2109–2112, 1989. URL `ftp://sig.enst.fr/pub/jfc/Papers/icassp89.ps.gz`.

C.K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541, 1994.

D. Casal. Decomposing autumn, 2008a.

D.P. Casal. Time after time: Short-circuiting the emotional distance between algorithm and human improvisors. In *proceedings of the International Computer Music Conference*, pages 24–29, 2008b.

D.P. Casal and M. Casey. Decomposing autumn: A component-wise recomposition. In *ICMC*, 2010.

M. Casey. Soundspotting: A new kind of process. *The Oxford Handbook of Computer Music*, 2009.

M. Casey. http://soundspotter.org/, May 2011.

M.A. Casey and A. Westner. Separation of mixed audio sources by independent subspace analysis. In *Proceedings of the International Computer Music Conference*, pages 154–161, 2000.

M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: current directions and future challenges. *Proceedings of the IEEE*, 96(4):668–696, 2008.

Michael Casey. http://digitalmusics.dartmouth.edu/ mcasey/bregman.

P. Chordia and A. Rae. Using source separation to improve tempo detection. In *Proceedings of the International Society for Music Information Reitreval (ISMIR)*, Kobe, JP, 2009.

N. Collins. Beyond notation: Communicating music. *Leonardo Music Journal*, pages 5–6, 2011.

S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.

I. Deliege. Grouping conditions in listening to music: An approach to lerdahl & jackendoff's grouping preference rules. *Music perception*, pages 325–359, 1987.

D. Deutsch. Grouping mechanisms in music. *The Psychology of Music*, 2:299–348, 1982.

Ringo Doe. Ishkur electronic music guide, October 2010. URL `http://techno.org/electronic-music-guide/`.

F. Doshi-Velez and Z. Ghahramani. Correlated non-parametric latent feature models. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pages 143–150. AUAI Press, 2009.

S.C. Douglas, A. Cichocki, and S.I. Amari. Multichannel blind separation and deconvolution of sources with arbitrary distributions. In *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop*, pages 436–445. IEEE, 1997.

J.S. Downie. The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology*, 29(4):247–255, 2008.

H Dufourt. Musique spectrale. *Musique, pouvoir, écriture*, pages 289–290, 1979.

M. Elhilali and S. Shamma. A biologically-inspired approach to the cocktail party problem. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pages V–V. IEEE, 2006.

R. Erickson. *Sound structure in music*. Univ of California Pr, 1975.

P.J. Fasciano, C.A. Rawley, T.R. Hegg, M. Leathurby, J.L. Bedell, and J.A. Ravan Jr. Digital audio workstations providing digital storage and display of video information, 1995. US Patent 5,467,288.

J. Foote et al. Content-based retrieval of music and audio. In *Multimedia Storage and Archiving Systems II, Proc. of SPIE*, volume 3229, pages 138–147, 1997.

E.B. Fox, E.B. Sudderth, M.I. Jordan, and A.S. Willsky. The sticky hdp-hmm: Bayesian nonparametric hidden markov models with persistent states. *Arxiv preprint arXiv*, 905, 2009.

E.B. George and M.J.T. Smith. Analysis-by-synthesis/overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones. *J. Audio Eng. Soc*, 40(6):497–516, 1992.

O. Gillet and G. Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3):529–540, 2008. ISSN 1558-7916.

G.H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14(5):403–420, 1970.

M.J. Grant. *Serial music, serial aesthetics: compositional theory in post-war Europe*, volume 16. Cambridge Univ Press, 2005.

T. Griffiths and Z. Ghahramani. Infinite latent feature models and the indian buffet process. *Advances in neural information processing systems*, 18:475, 2006.

G. Grisey. *Écrits ou L'invention de la musique spectrale*. MF, 2008.

Gérard Grisey. *Partiels Pour 18 Musiciense*. Ricordi, 1975.

R. Hennequin, B. David, and R. Badeau. Scale-invariant probabilistic latent component analysis. Technical report, Telecom ParisTech, Tech. Rep, 2011.

M.D. Hoffman. *Probabilistic Graphical Models for the Analysis and Synthesis of Musical Audio*. Princeton University, 2010.

M.D. Hoffman, P. EDU, D.M. Blei, and P.R. Cook. Bayesian nonparametric matrix factorization for recorded music. Citeseer, 2010.

T. Hoffman. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 50–57. ACM Press, New York, 1999.

T.S. Hughes. Groove and flow: Six analytical essays on the music of stevie wonder. *PhD diss., University of Washington*, page 3898A, 2003.

D. Ingram. " the clutter of the unkempt forest": John cage, music and american environmental thought. *Amerikastudien/American Studies*, pages 567–579, 2006.

G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 538–543. ACM, 2002.

T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *Communication Technology, IEEE Transactions on*, 15(1):52–60, 1967.

S. Kim-Cohen. *In the Blink of an ear:Toward a Non-Cochlear Sonic Art*. Contiuum, 2009.

J.F.C. Kingman. *Poisson processes*. Wiley Online Library, 1993.

M. Klingbeil. Software for spectral analysis, editing, and synthesis. In *Proceedings of the International Computer Music Conference*, pages 107–110. Citeseer, 2005.

C.L. Krumhansl. *Internal representations for music perception and performance*. Number 1557981272. American Psychological Association, 1992.

S.J. Kung, O.J.L. Tzeng, D.L. Hung, and D.H. Wu. Dynamic allocation of attention to metrical and grouping accents in rhythmic sequences. *Experimental Brain Research*, pages 1–14, 2011.

D. LaBerge. Attentional processing in music listening: A cognitive neuroscience approach. *Psychomusicology: Music, Mind and Brain*, 14(1-2):20–34, 2010.

S. Lakatos, S. McAdams, and R. Caussé. The representation of auditory source characteristics: Simple geometric form. *Attention, Perception, & Psychophysics*, 59(8):1180–1190, 1997.

D.D. Lee, H.S. Seung, et al. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

F. Lerdahl, R. Jackendoff, and R.S. Jackendoff. *A generative theory of tonal music*. The MIT Press, 1996. ISBN 026262107X.

S. Atkinson M. Casey. Strange charmed, 1999.

C.K. Madsen, R.V. Brittin, and D.A. Capperella-Sheldon. An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education*, 41(1):57–69, 1993.

A.J. Marcel. Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive psychology*, 15(2):197–237, 1983.

A. McGuiness and K. Overy. Music, consciousness, and the brain: music as shared experience of an embodied present. *Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives*, page 245, 2011.

M.E. McIntyre, R.T. Schumacher, and J. Woodhouse. On the oscillations of musical instruments. *The Journal of the Acoustical Society of America*, 74:1325, 1983.

P. Mermelstein. Distance measures for speech recognition, psychological and instrumental. *Pattern Recognition and Artificial Intelligence*, 116, 1976.

T.K. Moon. The expectation-maximization algorithm. *Signal Processing Magazine, IEEE*, 13(6):47–60, 1996.

V. Moscovich. French spectral music: An introduction. *Tempo (New Series)*, 3 (200):21–28, 1997.

T. Murail. *Désintégrations*. Éditions Salabert, 1989.

T. Murial. *Godwana*. Transatlantiques, 1980.

H. Okamoto, H. Stracke, W. Stoll, and C. Pantev. Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. *Proceedings of the National Academy of Sciences*, 107(3):1207, 2010.

I. Orife, I. Fred, O. Orife, et al. Riddim: A rhythm analysis and decomposition tool based on independent subspace analysis. 2001.

P. Osentinsky. Stratovinsky, 2010.

P. Osetinsky. Stratosound. Master's thesis, Dartmouth College, 2010.

J. Perry. Cage's sonatas and interludes for prepared piano: Performance, hearing and analysis. *Music Theory Spectrum*, 27(1):35–66, 2005.

MD Plumbley, SA Abdallah, JP Bello, ME Davies, and G. Monti. Submitted for publication in cybernetics and systems. prepared with the t&f journal template. automatic music transcription and audio source separation. 2007.

K. Qingyuan, S. Sarroff, S. Topel, and M.A. Casey. Getting into the groove with probabilistic latent component analysis. In *Neural Information Processing Systems (NIPS)*, Sierra-Nevada, Spain, 2011.

B. Raj and P. Smaragdis. Latent variable decomposition of spectrograms for single channel speaker separation. In *Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on*, pages 17–20. IEEE, 2005.

C. Rhodes, M. Casey, S. Abdallah, and M. Sandler. A markov-chain monte-carlo approach to musical audio segmentation. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pages V–V. IEEE, 2006.

J.C. Risset and D.L. Wessel. Exploration of timbre by analysis and synthesis. *The psychology of music*, pages 26–58, 1982.

P. Roland and J.S. Downie. Recent developments in the music encoding initiative project: Enhancing digital musicology and scholarship. In *Proc. 19th Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (Digital Humanities 2007)*, pages 186–189, 2007.

D.F. Rosenthal and H.G. Okuno. *Computational auditory scene analysis*. CRC, 1998.

A. Sa'id. http://www.beattips.com/beattips/amir-said/, September 2011.

P. Schaeffer. Solfege de l'objet sonore. *INA/GRM*, 1967.

Eric D. Scheirer. Bregman's chimerae: Music perception as auditory scene analysis. In *Proceedings of the International Conference on Music Perception and Cognition*, Montreal, CA, 1996.

K.R. Scherer and J.S. Oshinsky. Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, 1(4):331–346, 1977.

D. Schwartz. *Data-driven concatenative sound synthesis*. PhD thesis, University of Paris, 2004.

S. Serafini. Timbre judgments of javanese gamelan instruments by trained and untrained adults. *Psychomusicology: Music, Mind and Brain*, 14(1-2):137–153, 2010.

Y. Shao and D.L. Wang. Robust speaker identification using auditory features and computational auditory scene analysis. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 1589–1592. IEEE, 2008.

M. Shashanka. *Latent variable framework for modeling and separating single-channel acoustic sources*. PhD thesis, University Microfilms International, P. O. Box 1764, Ann Arbor, MI, 48106, USA, 2008.

Madhusudana Shashanka, Bhiksha Raj, and Paris Smaragdis. Probabilistic latent variable models as non-negative factorizations. Technical report, Mitsubishi Electric Research Laboratories, December 2007. URL `http://www.merl.com/reports/docs/TR2007-083.pdf`.

M. Slaney. A critique of pure. *Computational auditory scene analysis*, page 27, 1998.

D. Smalley. Spectromorphology: explaining sound-shapes. *Organised Sound*, 2 (2):107–126, 1997.

P. Smaragdis and J.C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 177–180. IEEE, 2003.

Paris Smaragdis, Bhiksha Raj, and Madhusudana V. S. Shashanka. Sparse and shift-invariant feature extraction from non-negative data. In *ICASSP*, pages 2069–2072. IEEE, 2008. ISBN 1-4244-1484-9. URL `http://dx.doi.org/10.1109/ICASSP.2008.4518048`.

J.O. Smith. Mathematics of the discrete fourier transform (dft). *URL: http://crma. standford. edu/-jos/mdft/Round Off Error Variance. html*, 2003.

I. Stravinsky. *Rite of Spring*. Dover Publications, 1989.

Y.W. Teh. Dirichlet process. *submitted to Encyclopedia of Machine Learning*, 2007.

Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.

Y.W. Teh, D. Gorur, and Z. Ghahramani. Stick-breaking construction for the indian buffet process. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, volume 11. Citeseer, 2007.

L. Thoresen and A. Hedman. Spectromorphological analysis of sound objects: an adaptation of pierre schaeffer's typomorphology. *Organised Sound*, 12(2): 129–141, 2007.

S. Topel. Violine for solo violin and live-electronics, 2010.

S. Topel. Elementary sources for string quartet and live-electronics, 2011a.

S. Topel. Intersecare for solo cello and live-electronics, 2011b.

S. Topel. Dissertation matlab code examples, May 2012a. URL `http://digitalmusics.dartmouth.edu/~spencer/dissC.zip`.

S. Topel. Dissertation music examples, audio and score files, May 2012b. URL `http://digitalmusics.dartmouth.edu/~spencer/dissM.zip`.

S. Topel and M. Casey. Timbre-based percussive rhythm classification and retrieval. In *Proceedings of the International Society for Music Information Reitreval (ISMIR)*, Utrecht, NL, 2011a.

S. Topel and M.A. Casey. Elementary sources: Latent component analysis for music composition. In *Proceedings of the International Society for Music Information Reitreval (ISMIR)*, Miami, USA, 2011b.

E. Tsunoo, N. Ono, and S. Sagayama. Musical bass-line pattern clustering and its application to audio genre classification. In *Proc. of ISMIR*, 2009.

M.E. Veal. *Dub: soundscapes and shattered songs in Jamaican reggae*. Wesleyan Univ Pr, 2007.

W.N. Venables and B.D. Ripley. *Modern applied statistics with S*. Springer verlag, 2002.

N. Virag. Single channel speech enhancement based on masking properties of the human auditory system. *Speech and Audio Processing, IEEE Transactions on*, 7(2):126–137, 1999.

T. Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3):1066–1074, 2007. ISSN 1558-7916.

D.L. Wang. On ideal binary mask as the computational goal of auditory scene analysis. *Speech separation by humans and machines*, pages 181–197, 2005.

D.L. Wang and G.J. Brown. *Computational auditory scene analysis: Principles, algorithms, and applications*. IEEE Press, 2006.

L. Wedin and G. Goude. Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology*, 13(1):228–240, 1972.

R. Weiss and J. Bello. Identifying repeated patterns in music using sparse convolutive non-negative matrix factorization. In *Proceedings of the International Society for Music Information Reitreval (ISMIR)*, pages 123–128, Utrecht, NL, 2010.

R. Weiss, J. Bello, et al. Shift-invariant probabilistic latent component analysis. 2010.

M. Wertheimer. Laws of organization in perceptual forms. *A source book of Gestalt psychology*, pages 71–88, 1938.

The Winstons. Amen brother, 1969.

T. Wishart and S. Emmerson. On sonic art. *12*, Routledge, 1996.

J. Woodruff, B. Pardo, and R. Dannenberg. Remixing stereo music with score-informed source separation. In *Proc. ISMIR*, pages 314–319. Citeseer, 2006.

B.R. Zendel and C. Alain. Concurrent sound segregation is enhanced in musicians. *Journal of cognitive neuroscience*, 21(8):1488–1498, 2009.